Unbiased Estimation of Reliability in Stress-Strength Model and Similar Problems

MIKHAIL S. NIKULIN

Steklov Mathematical Institute, 27 Fontanka, 191011, Saint Petersburg, Russia

VASSILY G. VOINOV

Kazakhstan Institute of Management, Economics and Strategic Research, Almaty, Kazakhstan

Abstract: Some problems related to an application of the unbiased estimators for sampling inspection, lifetime and reliability testing are discussed. Continuous and discrete, univariate and multivariate parametric probability models are considered.

Keywords and phrases: unbiased estimation, life distributions, multivariate failure rates, multivariate polynomial laws, sampling inspection, reliability and quality of products, stress-strength model

1. Introduction

Problems of estimating probabilities like $\mathbf{P}(\mathbf{X} < \mathbf{Y})$, $\mathbf{P}(\mathbf{X} < \mathbf{a})$, etc., $\mathbf{X}$ and $\mathbf{Y}$ being random vectors and $\mathbf{a}$ being a given threshold vector, are often encountered in many applications of statistics. Such problems arise most often in assessing the lifetime and reliability (Basu (1988), Voinov and Nikulin (1993, 1996)) and testing the quality of products (Belyaev(1975)). Among other applications it is worthy to mention a probabilistic description of radioactively polluted lands (Nikulin et al(1999)), sampling surveying of populations (Nikulin, Price, Turetayev, Voinov(1999)) and problems of texture satellite images recognition.

The problem of estimating the reliability of elements and systems is one of the main problems in engineering. Suppose that some stress is applied to an element or to a whole product. A product is destroyed if at some moment of time the stress applied exceeds the strength limit of it. The stress is a function of the equipment on which this element or system is tested and its value at each moment of time can be described by a random variable. The strength limit of the product is defined by the stress at which destruction occurs. It depends on material properties, technology, etc. If a few randomly chosen products being in mass production are tested then their strength limit x can also be described by a random variable X independent on Y describing the stress. The strength limit x and the maximum stress y are realizations of the random variables X and Y, respectively, and therefore the probability $\mathbf{P}\{Y < X\}$ that the maximum stress does not exceed the strength limit during the testing time $t \in [0, T]$ is considered to be a system reliability (see, e.g., Chandra and Owen (1975)). Naturally, products are not destroyed in every test but since this takes place, it is necessary to be able to get good estimators of the probability $\mathbf{P}\{Y < X\}$. Sometimes a supposition that components of a system fail independently of each other is inappropriate (see e.g. Lindley and Singpurwalla (1986)). In such cases one should consider multivariate models with correlated components. In this case probabilities of interest look like $\mathbf{P}(\mathbf{Y} < \mathbf{X})$, where $\mathbf{Y}$ and $\mathbf{X}$ are random vectors. Different techniques are used for estimating such probabilities. Maximum likelihood estimators as well as method of moments estimators seem to be the most widely used. Nonparametric estimators also have interesting applications (see, e.g., Bakker (1990), Van der Laan (1991,1994)). Nevertheless, if there exist the complete sufficient statistics and parametric functions of interest are estimable, one may use the minimum variance unbiased estimators. It is important, first, because reliability tests are rather often made for several samples obtained independently and, based on these samples, a combined estimator (usually the arithmetic mean one) is constructed. If the estimators used are biased, the averaged estimator may have a bias greater than its standard error and, hence, may cause overestimation or underestimation of the reliability of products. Secondly, a biased estimator can inadmissibly distort the information on the reliability of the system consisting of several stochastically independent subsystems, since the reliability estimator of the whole system is equal to the product of the reliabilities of its subsystems (Pugh (1963)). Many papers are devoted to the problem of reliability estimation in the case where random variables are independent identically or nonidentically distributed (see, e.g., Arnold (1975), Johnson (1988), Selvavel (1989), McCool (1991)).

For an inspection of products a random sampling is usually used. In view of this, a selection of an appropriate probability model is very important. Continuous and discrete, univariate and multivariate parametric models are frequently used. If measurements are of high precision, one may use continuous models. Sometimes one is unable to measure features of products with high accuracy. In such situations it is much better to quantize the measurements and use discrete probability models. To assess the quality of products for both continuous and discrete models, probabilistic measures like $\mathbf{P}(\mathbf{X} < \mathbf{a})$ or $\mathbf{P}(\mathbf{X} > \mathbf{a})$ are usually used. In this expression $\mathbf{X}$ stands for a random vector describing properties of a product and $\mathbf{a}$ is a prescribed threshold vector. Unbiased estimators of such probabilities for different univariate and multivariate, discrete and continuous models can be found in Voinov and Nikulin (1993,1996). That is why, in this paper we shall consider only a multivariate discrete probability model which relates to those probabilities and proves to be also useful in radioecology and sampling surveying.

2. Univariate continuous probability models

Let $\mathbb{X} = (X_1, ..., X_n)$ and $\mathbb{Y} = (Y_1, ..., Y_m)$ be samples, where

$$X \sim f(x;\theta),\ \theta \in \Theta,\ ,Y \sim g(y;\lambda),\ \lambda \in \Lambda.$$

In this case

$$\mathbf{P}\{Y < X\} = F(\theta, \lambda).$$

If the parameters θ and λ are unknown, the problem of the statistical estimation of the function $F(\theta, \lambda)$ arises.

Let, for example, $X_i \sim \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$, $i = 1, 2, ..., n$, and $Y \sim \varphi(y)$, i.e. the parameters of the maximum stress Y distribution are known,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

In this case $\mathbf{P}\{Y < X\} = \Phi(\mu/\sqrt{1+\sigma^2})$. The problem is to construct the best unbiased estimator $\hat{P}_1$ of the function $\Phi(\mu/\sqrt{1+\sigma^2})$. Using the Rao-Blackwell-Kolmogorov theorem, Mazumdar (1970) and Downton (1973) obtained this estimator for the case $n \geq 3$ as:

$$\hat{P}_1 = \frac{1}{B\left(\frac{1}{2}, \frac{n-2}{2}\right)} \int_{-1}^{1} \Phi\left(\bar{X}_n + \frac{uS_n(n-1)}{\sqrt{n}}\right)(1-u^2)^{\frac{n-4}{2}} du,$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ are statistics sufficient for μ and σ^2.

Solving the unbiasedness equation for the estimator of the function $\mathbf{P}\{Y < X\} = \Phi(\mu/\sqrt{1+\sigma^2})$, it can be shown (Voinov (1984)) that there exists the alternative expression for $\hat{P}_1$ being valid for $n \geq 2$ and equal to

$$\hat{P}_1 = \frac{1}{2} + \frac{1}{2\sqrt{\pi}} \sum_{m=0}^{\infty} \frac{(-1)^m (\bar{X}_n/\sqrt{2})^{2m+1}}{m!(m+1/2)} \times$$

$$\times {}_2F_1\left(-m, -m-\frac{1}{2}; \frac{n-1}{2}; \frac{(n-1)^2 S_n^2}{n\bar{X}_n}\right),$$

where ${}_2F_1(\alpha, \beta; \gamma; x)$ is the Gauss hypergeometrical function.

Let now parameters of the distribution of Y be unknown and

$$X_i \sim \frac{1}{\sigma_x}\varphi\left(\frac{x-\mu_x}{\sigma_x}\right), \; Y_i \sim \frac{1}{\sigma_y}\varphi\left(\frac{y-\mu_y}{\sigma_y}\right).$$

In this case

$$\mathbf{P}\{Y < X\} = \Phi\left(\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}\right).$$

The problem of estimating the probability $\mathbf{P}\{Y < X\}$, based on samples $\boldsymbol{X} = (X_1, ..., X_{n_1})$ and $\boldsymbol{Y} = (Y_1, ..., Y_{n_2})$ is much more realistic than the previous one where the parameters of the distribution of Y were supposed to be known. In Voinov (1984) it has been shown that for $n_1 \geq 3$ and $n_2 \geq 2$ the uniformly minimum variance unbiased estimator (UMVUE) $\hat{P}_2$ of the function $\Phi\left((\mu_x - \mu_y)/\sqrt{\sigma_x^2 + \sigma_y^2}\right)$ is

$$\hat{P}_2 = \begin{cases} 1 - \int_{-1}^{1} \psi(\rho_1)(1-u^2)^{\frac{n_1-4}{2}} du, & a > 1, \\ \int_{-1}^{1} \psi(\rho_2)(1-u^2)^{\frac{n_1-4}{2}} du & a < -1, \\ \frac{1}{2} + \frac{\Gamma\left(\frac{n_1-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{n_1-2}{2}\right)} \int_{-a}^{a} (1-u^2)^{\frac{n_1-4}{2}} du & \\ + \int_{a}^{1} \psi(\rho_3)(1-u^2)^{\frac{n_1-4}{2}} du - \int_{a}^{1} \psi(\rho_1)(1-u^2)^{\frac{n_1-4}{2}} du & |a| \leq 1, \end{cases}$$

where $\rho_1 = b(a+u)$, $\rho_2 = -\rho_1$, $\rho_3 = b(u-a)$,

$$a = \frac{\sqrt{n_1}(\bar{X} - \bar{Y})}{S(x)(n_1 - 1)}, \; b = \frac{\sqrt{n_2}(n_1-1)S(x)}{\sqrt{n_1}(n_2-1)S(y)},$$

$$S^2(x) = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2, \; S^2(y) = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2,$$

$$\psi(\rho_i) = \begin{cases} \frac{\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)\rho_i(1-\rho_i)^{\frac{n_2-2}{2}}}{2\pi\Gamma\left(\frac{n_1-2}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} & \\ \times{}_2F_1\left(1, \frac{n_2-1}{2}; \frac{n_2}{2}; 1-\rho_i^2\right), & 0 < \rho_i \le 1, \\ 0, & \rho_i > 1, \ i = 1, 2, 3. \end{cases}$$

Consider the problem of determining the reliability of a system consisting of N elements with different strength limits, which in tests is subjected to a stress during the time T. The system reliability is determined as the probability $q = \mathbf{P}(X < Y_1, X < Y_2, ..., X < Y_N)$, where random variables $Y_1, ..., Y_N$ correspond to strength limits and X is a random stress. Let all variables be stochastically independent, with $X \sim \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$, and Y_i $(i = 1, ..., N)$ having the p.d.f.

$$f(y;\alpha) = \begin{cases} \alpha e^{-\alpha y}, & y \ge 0, \ \alpha > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The parameter α is considered to be known. In this case, the probability q is equal (Chandra and Owen (1975)) to $q_1 + q_2$, where

$$q_1 = \Phi(-\mu/\sigma) \text{ and } q_2 = \exp\left\{-\alpha N\mu + \frac{\alpha^2 N^2\sigma^2}{2}\right\}\left[1 + \Phi\left(\alpha N\sigma - \frac{\mu}{\sigma}\right)\right].$$

If the parameter σ^2 is known, then, based on a sample $\boldsymbol{X} = (X_1, ..., X_n)$, one can construct the UMVUE of the function q, being equal to $\hat{q}(\sigma) = \hat{q}_1(\sigma) + \hat{q}_2(\sigma)$, where (see Chandra and Owen (1975)) $\hat{q}_1(\sigma) = \Phi(-\sqrt{n}\bar{X}_n/\sigma\sqrt{n-1})$,

$$\bar{q}_2(\sigma) = \exp\{-\alpha N\bar{X}_n + \alpha^2 N^2(n-1)\sigma^2/2n\}\times$$
$$\times\left[1 - \Phi\left(\frac{\sqrt{n-1}\alpha\sigma N}{\sqrt{n}} - \frac{\sqrt{n}\bar{X}_n}{\sigma\sqrt{n-1}}\right)\right].$$

Consider a more typical case when the parameter σ^2 is unknown. As has been shown by Lieberman and Resnikoff (1955) the UMVUE $\hat{q}_1$ of the function $q_1 = \Phi(-\mu/\sigma)$ is equal to

$$\hat{q}_1 = \begin{cases} 0, & z < 0, \\ I_z\left(\frac{n-2}{2}, \frac{n-2}{2}\right), & 0 \le z \le 1, \\ 1, & z > 1, \end{cases}$$

where

$$z = \frac{1}{2}\left(1 - \frac{\bar{X}_n\sqrt{n}}{(n-1)S_n}\right)$$

and $I_z(\alpha, \beta)$ is the incomplete β-function. To derive the UMVUE of the function $\hat{q}_2(\sigma)$ with respect to the unknown parameter σ^2, we represent $\hat{q}_2(\sigma)$ as $g(\sigma)\exp\{-\alpha N\bar{X}_n\}$, where

$$g(\sigma) = e^{\frac{\alpha^2N^2(n-1)\sigma^2}{2n}}\left[1 - \Phi\left(\frac{\sqrt{n-1}\alpha\sigma N}{\sqrt{n}} - \frac{\sqrt{n}\bar{X}_n}{\sigma\sqrt{n-1}}\right)\right]$$
$$= e^{\frac{\alpha^2N^2(n-1)\sigma^2}{2n}} - \frac{1}{\sigma}\int_{-\frac{\sqrt{n}\bar{X}_n}{\sqrt{n-1}}}^{\infty}\varphi\left(\frac{y}{\sigma}\right)\exp\left(-\frac{\sqrt{n-1}\alpha N}{\sqrt{n}}y\right)dy.$$

According to the known result of Neyman and Scott (1960) the UMVUE T_1 of the function $\exp\left(\frac{\alpha^2 N^2 (n-1)\sigma^2}{2n}\right)$ is the statistic

$$T_1 = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\left[\frac{T(n-1)^2\alpha^2 N^2}{4n}\right]^{\frac{n-3}{4}}} I_{\frac{n-3}{2}}\left(\frac{\alpha N(n-1)S_n}{\sqrt{n}}\right).$$

Solving the equation of unbiasedness $ET_2 = \frac{1}{\sigma}\varphi(\frac{y}{\sigma})$ w.r.t. T_2, it can be shown (Voinov and Nikulin (1993)) that the UMVUE $\hat{q}$ of q is $\hat{q}_1 + T_3$, where

$$T_3 = \begin{cases} T_1 e^{-\alpha N \bar{X}_n}, & \bar{X}_n \le -\frac{S_n(n-1)}{\sqrt{n}} \\ T_1 e^{-\alpha N \bar{X}_n} - \frac{\Gamma\left(\frac{n-1}{2}\right)e^{-\alpha N \bar{X}_n}}{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)} \int\limits_{-\frac{\sqrt{n}\bar{X}_n}{S_n(n-1)}}^{1} (1-x^2)^{\frac{n-4}{2}} & \\ \times \exp\left\{-\frac{\alpha N S_n(n-1)x}{\sqrt{n}}\right\} dx, & |\bar{X}_n| < \frac{S_n(n-1)}{\sqrt{n}}, \\ 0, & \bar{X}_n \ge \frac{S_n(n-1)}{\sqrt{n}}. \end{cases}$$

Let $\boldsymbol{X} = (X_1, X_2, ..., X_m)^T$ and $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)^T$ be two independent random samples from families with probability density functions $f_1(x;\theta_1)$ and $f_2(y;\theta_2)$ respectively, where

$$f_1(x;\theta_1) = q_1(\theta_1)h_1(x), \qquad a < x < \theta_1 < b, \tag{1}$$

and

$$f_2(y;\theta_2) = q_2(\theta_2)h_2(y), \qquad a < y < \theta_2 < b. \tag{2}$$

It is supposed that $-\infty < a < b < \infty$ are known constants; $h_1(x), h_2(y)$ are positive continuous functions and $q_1(\theta_1), q_2(\theta_2)$ are differentiable, where

$$\frac{1}{q_i(\theta)} = \int\limits_a^\theta h_i(x)dx, \qquad i = 1, 2.$$

Let $X_{(1)} < X_{(2)} < ... < X_{(m)}$ and $Y_{(1)} < Y_{(2)} < ... < Y_{(n)}$ be sets of order statistics from (**??**) and (**??**). The complete sufficient statistics for parameters θ_1 and θ_2 are known to be $X_{(m)}$ and $X_{(n)}$ respectively. Rohatgi (1989) solving the unbiasedness equation showed that the UMVUE $\varphi(X_{(m)}, Y_{(n)})$ of an estimable function $g(\theta_1, \theta_2)$ is

$$\begin{aligned} \varphi(X_{(m)}, Y_{(n)}) = g(X_{(m)}, Y_{(n)}) + \tfrac{g_1(X_{(m)}, Y_{(n)})}{m q_1(X_{(m)}) h_1(X_{(m)})} + \\ + \tfrac{g_2(X_{(m)}, Y_{(n)})}{n q_2(Y_{(n)}) h_2(Y_{(n)})} + \tfrac{g_{12}(X_{(m)}, Y_{(n)})}{mn q_1(X_{(m)}) q_2(Y_{(n)}) h_1(X_{(m)}) h_2(Y_{(n)})}, \end{aligned} \tag{3}$$

where

$$g_1(x,y) = \frac{\partial g(x,y)}{\partial x}, \quad g_2(x,y) = \frac{\partial g(x,y)}{\partial y}, \quad g_{12}(x,y) = \frac{\partial^2 g(x,y)}{\partial x \partial y}.$$

Suppose the samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ are Type II censored. Another words one observes only $X_{(1)}, X_{(2)}, ..., X_{(r_1)}$, $1 \le r_1 \le m$, and $Y_{(1)}, Y_{(2)}, ..., Y_{(r_2)}$, $1 \le r_2 \le n$, respectively. In this case the UMVUE $\varphi(X_{(r_1)}, Y_{(r_2)})$ of an estimable function $g(\theta_1, \theta_2)$ is (Selvavel (1989))

$$\begin{aligned}\varphi(X_{(r_1)}, Y_{(r_2)}) = \tfrac{(r_1-1)!(r_2-1)!}{m!n!} q_1^{r_1-1}(X_{(r_1)}) q_2^{r_2-1}(Y_{(r_2)}) \times \\ \times \mathcal{D}_{11}^{m-r_1+1} \left\{ \tfrac{1}{q_1^m(X_{(r_1)})} \mathcal{D}_{01}^{n-m-r_2+r_1} \left(\tfrac{g(X_{(r_1)}, Y_{(r_2)})}{q_2^n(Y_{(r_2)})} \right) \right\},\end{aligned} \tag{4}$$

if $n - m - r_2 + r_1 \geq 0$, and

$$\begin{aligned}\varphi(X_{(r_1)}, Y_{(r_2)}) = \tfrac{(r_1-1)!(r_2-1)!}{m!n!} q_1^{r_1-1}(X_{(r_1)}) q_2^{r_2-1}(Y_{(r_2)}) \times \\ \times \mathcal{D}_{11}^{m-r_2+1} \left\{ \tfrac{1}{q_2^n(Y_{(r_2)})} \mathcal{D}_{10}^{m-n-r_1+r_2} \left(\tfrac{g(X_{(r_1)}, Y_{(r_2)})}{q_1^m(X_{(r_1)})} \right) \right\},\end{aligned} \tag{5}$$

if $n - m - r_2 + r_1 < 0$, where

$$\mathcal{D}_{ij} = \mathcal{D}_{ij}\omega(\theta_1, \theta_2) = \frac{1}{h_1^i(\theta_1) h_2^j(\theta_2)} \frac{\partial^{i+j}\omega(\theta_1, \theta_2)}{\partial^i\theta_1 \partial^j\theta_2}.$$

Since

$$\mathbf{P}\{X < Y\} = \begin{cases} q_1(\theta_1)q_2(\theta_2) \int\limits_a^{\theta_2} \left(\int\limits_x^{\theta_2} h_2(y)dy \right) h_1(x)dx, & \theta_1 > \theta_2, \\ q_1(\theta_1)q_2(\theta_2) \int\limits_a^{\theta_1} \left(\int\limits_x^{\theta_2} h_2(y)dy \right) h_1(x)dx, & \theta_1 < \theta_2, \end{cases} \tag{6}$$

from (**??**), (**??**) and (**??**) it follows that the UMVUE $\psi(X_{r_1}, Y_{r_2})$ of $\mathbf{P}\{X < Y\}$ if exists is (Selvavel (1989))

$$\begin{aligned}\psi(X_{r_1}, Y_{r_2}) = \tfrac{(r_1-1)!(r_2-1)!}{m!n!} q_1^{r_1-1}(X_{(r_1)}) q_2^{r_2-1}(X_{(r_2)}) \times \\ \times \mathcal{D}_{11}^{m-r_1+1} \left[\tfrac{1}{q_1^m(X_{(r_1)})} \mathcal{D}_{01}^{n-m-r_2+r_1} \left(\tfrac{r(X_{(r_1)}, Y_{(r_2)})}{q_2^n(Y_{r_2})} \right) \right],\end{aligned} \tag{7}$$

if $X_{(r_1)} > Y_{(r_2)}$, and

$$\begin{aligned}\psi(X_{r_1}, Y_{r_2}) = \tfrac{(r_1-1)!(r_2-1)!}{m!n(m+r_2-r_1-1)} q_1^{r_1-1}(X_{(r_1)}) q_2^{r_2-1}(Y_{(r_2)}) \times \\ \times \mathcal{D}_{11}^{m-r_1+1} \left[\tfrac{1}{q_1^m(X_{(r_1)}) q_2^{m+r_2-r_1}(Y_{(r_2)})} \times \right. \\ \left. \times \left\{ r(X_{(r_1)}, Y_{(r_2)}) + \tfrac{(n-m-r_2+r_1)}{(m+r_2-r_1)} \right\} \right],\end{aligned} \tag{8}$$

if $X_{(r_1)} < Y_{(r_2)}$.

3. Multivariate continuous probability models

Let random variables X and Y follow the absolutely continuous bivariate exponential distribution of Block and Basu (1974). This distribution is widely used for describing both physical and biological systems.

The joint survival functions of this distribution is

$$\begin{aligned}\bar{F}(x, y) = \tfrac{\lambda}{\lambda_1+\lambda_2} \exp\{-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)\} - \\ - \tfrac{\lambda_{12}}{\lambda_1+\lambda_2} \exp\{-\lambda \max(x, y)\}, \qquad x, y > 0,\end{aligned} \tag{9}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$. From now on we shall follow the paper of Klein and Basu (1985), see also Basu(1988).

Let marginals of (**??**) be identical, i.e. $\lambda_1 = \lambda_2 = \alpha$. Denote $\lambda_{12} = \beta$. Let $\boldsymbol{X} = (\mathbf{X}_1, ..., \mathbf{X}_n)$ be a sample from (**??**) with $\mathbf{X}_i = (X_i, Y_i)^T, i = 1, 2, ..., n$. The complete sufficient for parameters λ and β statistics are

$$T = \sum_{i=1}^{n} \min(X_i, Y_i) \text{ and } V = \sum_{i=1}^{n} \max(X_i, Y_i) - \sum_{i=1}^{n} \min(X_i, Y_i).$$

The conditional density function of X_1 and Y_1 given $T = t$ and $V = v$ can be derived as

$$p(x_1, y_1 \mid t, v) = \begin{cases} \frac{(n-1)^2 (t-x_1)^{n-2} (v-y_1+x_1)^{n-2}}{2t^{n-1} v^{n-1}}, \\ \quad 0 < x_1 < t, \quad x_1 < y_1 < v + x_1; \\ \frac{(n-1)^2 (t-y_1)^{n-2} (v-x_1+y_1)^{n-2}}{2t^{n-1} v^{n-1}}, \\ \quad 0 < y_1 < t, \quad y_1 < x_1 < v + y_1. \end{cases} \tag{10}$$

Based on an unique observation $\mathbf{X}_1 = (X_1, Y_1)^T$ the statistic

$$\varphi(X_1, Y_1) = \begin{cases} 1, & X_1 > x, \quad Y_1 > y, \\ 0, & \quad \text{otherwise}, \end{cases} \tag{11}$$

is an unbiased estimator of $\bar{F}(x, y) = \mathbf{P}\{X > x, Y > y\}$. Hence, the conditional expectation of $\varphi(X_1, Y_1)$ given T and Y gives the UMVUE

$$\psi(T, V) = \mathbf{E}\{\varphi(X_1, Y_1) \mid T, V\}$$

of $\bar{F}(x, y)$. There are three different cases.

i) $T > x = y > 0$.

$$\psi(T, V) = \left(1 - \frac{x}{T}\right)^{n-1}.$$

ii) $x < y < T$.

$$\psi(T, V) = (1 - \frac{y}{T})^{n-1} + \frac{n-1}{2V^{n-1}T^{n-1}} \times$$

$$\times \sum_{k=0}^{n-1} (-1)^k \frac{\binom{n-1}{k} (V - y + T)^{n-k-1}}{n + k - 1} [(T - x)^{n+k-1} - (T - y)^{n+k-1}].$$

iii) $T > x > y > 0$.

$$\psi(T, V) = (1 - \frac{x}{T})^{n-1} + \frac{n-1}{2V^{n-1}T^{n-1}} \times$$

$$\times \sum_{k=0}^{n-1} (-1)^k \frac{\binom{n-1}{k} (V - x + T)^{n-k-1}}{n + k - 1} [(T - y)^{n+k-1} - (T - x)^{n+k-1}].$$

Let $\mathbf{X} = (X_1, ..., X_p)^T$ and $\mathbf{Y} = (Y_1, ..., Y_q)^T$ be two random vectors such that the distribution of $\mathbf{Z} = (\mathbf{X}^T, \mathbf{Y}^T)^T$ is multivariate normal $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\boldsymbol{\mu}_x^T, \boldsymbol{\mu}_y^T)^T = ((\mu_{x_1}, ..., \mu_{x_p})^T, (\mu_{y_1}, ..., \mu_{y_q})^T)^T$ and the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}$$

is $(p+q)\times(p+q)$ positive definite matrix.

Let $\mathbf{a}=(a_1,...,a_p)^T$ and $\mathbf{b}=(b_1,...,b_q)^T$ be two known vectors. The problem is to estimate $R=\mathbf{P}\{\mathbf{a}^T\mathbf{X}>\mathbf{b}^T\mathbf{Y}\}$.

This problem arises, e.g.,(see Gupta and Gupta (1990)) in systems where the energy is supplied to systems by p sources and is consumed through q sources. It is also useful in biochemistry of brain.

The distribution function of a random variable $U=\mathbf{b}^T\mathbf{y}-\mathbf{a}^T\mathbf{X}$ is $\Phi(\frac{u-\mu}{\sigma})$, where $\mu=\mathbf{b}^T\boldsymbol{\mu}_y-\mathbf{a}^T\boldsymbol{\mu}_x$ and $\sigma^2=\mathbf{a}^T\boldsymbol{\Sigma}_{xx}\mathbf{a}-2\mathbf{a}^T\boldsymbol{\Sigma}_{xy}\mathbf{b}+\mathbf{b}^T\boldsymbol{\Sigma}_{yy}\mathbf{b}$. From this it follows that

$$R=\mathbf{P}\{\mathbf{a}^T\mathbf{X}>\mathbf{b}^T\mathbf{Y}\}=\mathbf{P}\{U<0\}=\Phi(-\frac{\mu}{\sigma}).$$

Let $\mathbb{Z}=(\mathbf{Z}_1,...,\mathbf{Z}_n)$ be a random sample of size n from

$$\mathbf{Z}=(\mathbf{X}^T,\mathbf{Y}^T)^T\sim N_{p+q}(\boldsymbol{\mu},\boldsymbol{\Sigma}),$$

where $\mathbf{Z}_i=(\mathbf{X}_i^T,\mathbf{Y}_i^T)^T, i=1,2,...,n$.

The statistics $U_i=\mathbf{b}^T\mathbf{Y}_i-\mathbf{a}^T\mathbf{X}_i,\quad i=1,2,...,n$, form a random sample from $N(\mu,\sigma^2)$. Denoting

$$\bar U=\frac{1}{n}\sum_{i=1}^n U_i,\quad S^2=\frac{1}{n}\sum_{i=1}^n(U_i-\bar U)^2,\quad \bar{\mathbf{X}}=\frac{1}{n}\sum_{i=1}^n\mathbf{X}_i,$$

$$\mathbf{S_{xx}}=\frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i-\bar{\mathbf{X}})(\mathbf{X}_i-\bar{\mathbf{X}})^T,\quad \mathbf{S_{xy}}=\frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i-\bar{\mathbf{X}})(\mathbf{Y}_i-\bar{\mathbf{Y}})^T,$$

$$\mathbf{S_{yx}}=\frac{1}{n}\sum_{i=1}^n(\mathbf{Y}_i-\bar{\mathbf{Y}})(\mathbf{X}_i-\bar{\mathbf{X}})^T,\quad \mathbf{S_{yy}}=\frac{1}{n}\sum_{i=1}^n(\mathbf{Y}_i-\bar{\mathbf{Y}})(\mathbf{Y}_i-\bar{\mathbf{Y}})^T,$$

$$c=\frac{\bar U}{\sqrt{n-1}S}=\frac{\sqrt{n}(\mathbf{b}^T\bar{\mathbf{Y}}-\mathbf{a}^T\bar{\mathbf{X}})}{\sqrt{n-1}[\mathbf{a}^T\mathbf{S_{xx}}\mathbf{a}-2\mathbf{a}^T\mathbf{S_{xy}}\mathbf{b}+\mathbf{b}^T\mathbf{S_{yy}}\mathbf{b}]^{1/2}},$$

the UMVUE $\hat R$ of $R=\mathbf{P}\{\mathbf{a}^T\mathbf{X}>\mathbf{b}^T\mathbf{Y}\}$ can be written as (Gupta and Gupta (1990))

$$\hat R=\begin{cases}0, & c\geq 1,\\ \frac{1}{B\left(\frac{1}{2},\frac{n-2}{2}\right)}\int\limits_c^1(1-u^2)^{\frac{n-4}{2}}du, & |c|<1,\\ 1, & c\leq -1.\end{cases}$$

Multivariate discrete probability models

Consider a class of discrete distributions useful in assessing the joint survival function of a system with discrete time, sampling inspection, radioecology, etc. The class is induced by the following urn scheme. Let an urn contain balls marked by vectors $\mathbf{a}=(a_1,...,a_m)^T$, $a_j\in\{\tilde a_{j1},\tilde a_{j2},\ldots,\tilde a_{jk_j}\}$, $j=1,...,m$, $\tilde a_{ij}$ – being arbitrary integers. Denote by

$$\mathbf{a}_j=(a_{j1},a_{j2},\ldots,a_{jm})^T, j=1,2,\ldots,K,$$

where $K=k_1\cdot k_2\cdots k_m$, all possible values of $\mathbf{a}$ defined by possible values of their components a_j. Let $p_{\mathbf{a}_j}$, $j=1,2,\ldots,K$, be the probability for a ball bearing a vector $\mathbf{a}_j$ be drawn with replacement and $\sum\limits_{j=1}^K p_{\mathbf{a}_j}=1$.

Let a random vector $\mathbf{X} = (X_1, \ldots, X_m)^T$ take the value $\mathbf{r} = (r_1, \ldots, r_m)^T$ if sums of numbers of jth components of vectors on n balls drawn with replacement are

$$\sum_{i=1}^{K} a_{ij} l_{\mathbf{a}_i} = r_j,$$

$$n \min_{1 \le i \le k_j} \{\tilde{a}_{ji}\} \le r_j \le n \max_{1 \le i \le k_j} \{\tilde{a}_{ji}\}, \; j = 1, 2, ..., m,$$

where $l_{\mathbf{a}_i}, i = 1, \ldots, K$, denotes the number of balls in the sample bearing the vector $\mathbf{a}_i$ and $l_{\mathbf{a}_i}$th are nonnegative integers such that $\sum\limits_{i=1}^{K} l_{\mathbf{a}_i} = n$.

Evidently that

$$\mathbf{P}(\mathbf{X} = \mathbf{r}, \mathbf{p}) = \sum \frac{n!}{\prod\limits_{i=1}^{K} l_{\mathbf{a}_i}!} \prod_{i=1}^{K} p_{\mathbf{a}_i}^{l_{\mathbf{a}_i}}, \tag{12}$$

where $\mathbf{p} = (p_{\mathbf{a}_1}, \ldots, p_{\mathbf{a}_K})^T$ is the vector of parameters and the summation is performed over all sets of nonnegative solutions $l_{\mathbf{a}_i}, i = 1, 2, \ldots, K$, of the system of linear diophantine equations

$$\begin{cases} \sum\limits_{i=1}^{K} a_{ij} l_{\mathbf{a}_i} = r_j, \quad j = 1, 2, ..., m, \\ \sum\limits_{i=1} l_{\mathbf{a}_i} = n. \end{cases} \tag{13}$$

It is easily shown that (**??**) is a proper probability distribution because $\sum_{\mathbf{r}} \mathbf{P}(\mathbf{X} = \mathbf{r}, \mathbf{p}) = 1$. This distribution can be considered as a generalization of a probability distribution introduced in Panaretos and Xekalaki (1986).

To evaluate probabilities (**??**) one needs to find all solutions of the system of linear equations (**??**) in nonnegative integers or what is the same to construct all vector partitions of $\mathbf{r}$ satisfying the evident relation $\sum\limits_{i=1}^{K} l_{\mathbf{a}_i} = n$. Many algorithms solving a system of linear Diophantine equations are known. One may use, e.g., the Smith's technique. Unfortunately, this technique gives only the partial solution in integers. An algorithm giving the partial solution in integers in polynomial time is also known, but the problem of enumerating all solutions in nonnegative integers, which is NP-complete, still exists. Fortunately, using unbiased estimators for probabilities (**??**) it becomes possible to evaluate them exactly even for sample sizes of the order of $n = 16$ (Nikulin et al (1999).

Since $\sum\limits_{i=1}^{K} p_{\mathbf{a}_i} = 1$, probabilities (**??**) depend on $K - 1$ parameters. We may consider $\mathbf{P}(\mathbf{X} = \mathbf{r}, \mathbf{p})$ as the sum of multinomial probabilities

$$\mathbf{P}(\mathbf{Y} = \mathbf{l}, \mathbf{p}) = \frac{n!}{\prod\limits_{i=1}^{K} l_{\mathbf{a}_i}!} \prod_{i=1}^{K} p_{\mathbf{a}_i}^{l_{\mathbf{a}_i}}, \tag{14}$$

where $\mathbf{l} = (l_{\mathbf{a}_1}, \ldots, l_{\mathbf{a}_k})^T$. It is well known that the complete sufficient for parameters of (**??**) statistic is the subvector $\mathbf{Z}_{K-1} = (Z_1, \ldots, Z_{K-1})^T$ of the vector $\mathbf{Z} = (\mathbf{Z}_{K-1}, Z_K)^T$, where $Z_K = nM - \sum\limits_{i=1}^{K-1} Z_i$, M being the number of samples, each of which contains n

balls, and $Z_j = \sum_{i=1}^{M} X_{ij}, \quad j = 1, 2, \ldots, K-1$, where X_{ij} is the number of balls bearing $\mathbf{a}_j$ in the ith sample of n balls.

Using a standard technique (see, e.g., Voinov and Nikulin (1996)) it is easily proved that the UMVUE $\mathcal{P}_0 = \mathbf{P}(\mathbf{X} = \mathbf{r}|\mathbf{Z})$ of the probability $\mathbf{P}(\mathbf{X} = \mathbf{r}, \mathbf{p})$ is

$$\mathcal{P}_0 = \sum A(\mathbf{Z}, \mathbf{l}), \tag{15}$$

where

$$A(\mathbf{Z}, \mathbf{l}) = \begin{cases} \dfrac{\prod_{i=1}^{K} \binom{Z_i}{l_{\mathbf{a}_i}}}{\binom{nM}{n}}, & Z_i \geq l_{\mathbf{a}_i}, \qquad M \geq 1, \\ 0, & \text{otherwise}, \end{cases} \tag{16}$$

and the summation in (**??**) is performed over all sets of nonnegative solutions $l_{\mathbf{a}_i}$ of (**??**).

The UMVUE $\mathcal{P}_1$ of $\mathbf{Var}\mathcal{P}_0$ is

$$\mathcal{P}_1 = \mathcal{P}_0^2 - \mathcal{P}_2, \tag{17}$$

where

$$\mathcal{P}_2 = \sum B(\mathbf{Z}, \mathbf{l}) + 2 \sum_{i>j} C(\mathbf{Z}, \mathbf{l}), \tag{18}$$

$$B(\mathbf{Z}, \mathbf{l}) = \begin{cases} \dfrac{\prod_{k=1}^{K} \binom{2l_{\mathbf{a}_k}}{l_{\mathbf{a}_k}} \binom{Z_k}{2l_{\mathbf{a}_k}}}{\binom{2n}{n}\binom{nM}{2n}}, & Z_k \geq 2l_{\mathbf{a}_k}, \quad M \geq 2, \\ 0, & \text{otherwise}. \end{cases} \tag{19}$$

and

$$C(\mathbf{Z}, \mathbf{l}) = \begin{cases} \dfrac{\prod_{k=1}^{K} \binom{l^i_{\mathbf{a}_k} + l^j_{\mathbf{a}_k}}{l^i_{\mathbf{a}_k}} \binom{Z_k}{l^i_{\mathbf{a}_k} + l^j_{\mathbf{a}_k}}}{\binom{2n}{n}\binom{nM}{2n}}, & Z_k \geq l^i_{\mathbf{a}_k} + l^j_{\mathbf{a}_k}, \quad M \geq 2, \\ 0, & \text{otherwise}, \end{cases} \tag{20}$$

$\{l^i_{\mathbf{a}_1}, \ldots, l^i_{\mathbf{a}_K}\}$, $i = 1, 2, \ldots, q$, is the ith set of arbitrarily ordered sets of solutions of the system of equations (**??**), $1 \leq i, j \leq q$, where q is the total number of sets of solutions for $\mathbf{r}$ given. The first sum in (**??**) is performed over all q sets and the second one over sets with $i > j$.

Formulas (**??**) and (**??**) are easily proved by direct evaluation of their expectations with respect to the probability distribution of the complete sufficient for parameters $p_{\mathbf{a}_i}$ statistic $\mathbf{Z}_{K-1}$. It is worthy to note that UMVUEs $\mathcal{P}_0$ are probabilities themselves, i.e., $\sum_{\mathbf{r}} \mathcal{P}_0 = 1$. Another words, they are range preserving in the sense of Hoeffding (1983).

References

1. Basu A.P. (1988). Multivariate exponential distributions and their applications in reliability, *P.R.Krishnaiah and C.R.Rao (eds.), Handbook of Statistics*, **7**, 467–476.

2. Belyaev Yu.K. (1975).*Probabilistic methods of sampling inspection*, Nauka, Moscow (in Russian).

3. Block H.W., Basu A.P. (1974). A continuous bivariate exponential extension, *JASA*, **69**, 1031–1037.

4. Chandra S. and Owen D.B. (1975). On estimating the reliability of a component subject to several different stresses (strengths), *Naval Res. Logist. Quart.*, **22**, 31–39.

5. Downton F. (1973). The estimation of $\mathbf{P}\{Y < X\}$ in the normal case, *Technometrics*, **15**, 551–558.

6. Gupta R.D., Gupta R.C. (1990). Estimation of $Pr(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y})$ in the multivariate normal case, *Statistics*, **21**, 91–97.

7. Hoeffding W. (1983). Range-preserving unbiased estimators in the multinomial case, *Inst.Statist.Mimeo Ser.*, **1515**, 1–10.

8. Klein J.P., Basu A.P. (1985). Estimating reliability for bivariate exponential distributions, *Sankhyá*, **47B**, 346–353.

9. Lieberman G.J., Resnikoff G.J. (1955). Sampling plans for inspection by variables, *JASA*, **50**, 457–516.

10. Lindley D.V., Singpurwalla N.D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment, *J. Appl. Prob.*, **23**, 418–431.

11. Mazumdar M. (1983). Some estimates of reliability using interference theory, *Naval Res. Logist. Quart.*, **17**, 159–165.

12. Neyman J., Scott E.L. (1960). Correction for bias introduced by a transformation of variables, *Ann. Math. Stat.*, **31**, 645–656.

13. Nikulin M.S., Novak M.M., Smirnov T.I., Voinov V.G. (1999). A probabilistic description of radioactive contamination, (to be published in *Applied Radiation and Isotops*).

14. Nikulin M.S., Price P.M., Turetayev D., Voinov V.G. (1999). An application of a multivariate discrete probability distribution for sampling survey analysis of supply chains, *Technical Report N9905, University Victor Segalen Bordeaux 2, Bordeaux.*

15. Panaretos J., Xekalaki E. (1986). On generalized binomial and multinomial distributions and their relation to generalized Poisson distributions, *Ann.Inst.Statist.Math.*, **38A**, 223–231.

16. Pugh E.L. (1963). The best estimate of reliability in the exponential case, *Oper. Res.*, **11**, 57–61.

17. Rohatgi V.K. (1989). Unbiased estimation of parametric functions in sampling from one-truncation parameter families, *Austr. J. Stat.*.

18. Selvavel K. (1989). Unbiased estimation in sampling from two-truncation parameter families when both samples are type II censored, *Commun.Stat.:Theory and Meth.*, **18**, 3519–3531.

19. Unnikrishnan Nair N., Asha G. (1997). Some classes of multivariate life distributions in discrete time, *J. of Multivar. Anal.*, **62**, 181–189.

20. Voinov V.G. (1984). On unbiased estimation of $\mathbf{P}\{Y < X\}$ in the normal case, *Zap. Nauchn. Sem. LOMI*, **136**, 5–12(in russian).

21. Voinov V.G., Nikulin M.S. (1993). *Unbiased estimators and their applications. Vol.1:Univariate case*, Dordrecht/Boston/London: Kluwer.

22. Voinov V.G., Nikulin M.S. (1996). *Unbiased estimators and their applications. Vol.2:Multivariate case*, Dordrecht/Boston/London: Kluwer.