

## طراحی آماری نمودار کنترلی چندمتغیره ژرفاپایه مبتنی بر صدک

شادی نصراله زاده

دانشجوی دکتری، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، گروه آمار، تهران، ایران. sh\_nasrollahzadeh@azad.ac.ir

محمد بامنی مقدم

(نویسنده مسئول)، استاد، گروه آمار دانشکده‌ی علوم ریاضی، دانشگاه علامه طباطبائی، تهران، ایران. bamenimoghadam@atu.ac.ir

**چکیده:** با استفاده از رویکرد مبتنی بر صدک، روشی را برای طراحی آماری نمودار کنترلی ژرفاپایه معرفی می‌کنیم. آماره‌ی نمودار کنترلی پیشنهادی، ناوردای آفین و مجاناً آزاد توزیع است. معمولاً عملکرد نمودار کنترلی با معیار متوسط طول اجرا ارزیابی می‌شود. طول اجرا دارای توزیع هندسی و با انحراف استاندارد بزرگی، چوله به سمت راست است و شاید معیار مناسبی برای ارزیابی نمودار کنترلی نباشد. بنابراین از روش طراحی آماری نمودارهای کنترلی با رویکرد مبتنی بر صدک استفاده می‌کنیم که بهبود و توسعه‌ای بر طراحی آماری کلاسیک است. با قرار دادن شرایطی روی طول اجراها، با احتمال‌های از پیش تعیین شده طول اجراهای تحت کنترل و خارج از کنترل، تضمین می‌شوند و می‌توان مطمئن بود طول اجرای تحت کنترل بیش از مقدار دلخواه و طول اجرای خارج از کنترل کمتر از مقدار دلخواه است. شبیه‌سازی‌ها نشان می‌دهند نمودار کنترلی با رویکرد مبتنی بر صدک در مقایسه با رویکرد متوسط طول اجرا کارا تر است.

**واژگان کلیدی:** طول اجرا، طراحی آماری، کنترل فرایند آماری، نمودار کنترلی صدکی، مکان چند متغیره

بالا که داده‌های چندمتغیره را پایش می‌کنند، کاهش بُعد مسئله تا حد ممکن، بدون از دست دادن اطلاعات معنی‌دار، در رابطه با هشدار ارائه شده است. اندازه‌های مؤثر و کارایی که به مقدار کافی کاهش داده‌ها را برای مسائل مربوط به تعداد زیاد متغیر، در یک روش کاملاً ناپارامتری، فراهم کرده‌اند، بر پایه‌ی ژرفای داده‌ها هستند. به همین جهت اخیراً پژوهش‌های تحلیل چند متغیره ناپارامتری ژرفاپایه داده‌ها، که نیازی به فرضیه نرمال بودن و هیچ قیدی روی ابعاد داده‌ها ندارند، جذابیت بیشتری پیدا کرده است.

تکنیک‌های ژرفای داده‌ها برای کاهش بُعد، بر پایه‌ی اندازه‌گیری میزان مرکزیت یا دورافتادگی یک مشاهده چندمتغیره (یا نمونه اصلی) نسبت به توزیع یا جامعه تحت بررسی است. مرکزیت یا دورافتادگی یک نقطه مورد نظر در ابر داده‌ها اندازه‌گیری می‌شود، که ابر داده‌ها گروه مشاهدات مربوط به یک متغیر تصادفی چندمتغیره است. بنابراین رتبه‌بندی مرکز به بیرون نقاط داده‌ها می‌تواند با استفاده از نظریه ژرفای داده‌ها انجام شود. به طور آشکار، اگر مقادیر ژرفاها برای همه نقاط محاسبه و مقایسه شده باشند، ژرفای داده‌ها موجب ایجاد رتبه‌بندی مرکز به بیرون نقاط نمونه می‌شود. این نوع از رتبه‌بندی مرکز به بیرون

### ۱- مقدمه

در کنترل فرایند آماری چندمتغیره ( $MSPC^1$ )، نمودارهای کنترلی ابزار بسیار مهمی برای پایش و/یا کنترل فرایندهای تولید هستند. مطالعه نمودارهای کنترلی چندمتغیره اولین بار توسط هتلینگ (۱۹۴۷) انجام شد [۱]. سپس پژوهشگرهای دیگری در این زمینه مطالعاتی انجام دادند [۲]. نمودارهای کنترلی در نظر گرفته شده توسط این افراد بر اساس فرضیه نرمال بودن برای متغیرهای کیفیت است. در این ارتباط، نمودارهای کنترلی  $T^2$ - هتلینگ به‌طور گسترده برای پایش بردار مکان فرایند استفاده می‌شوند که در آن‌ها فرضیه‌های نرمال چندمتغیره بودن متغیرهای پاسخ (مشخصه‌های کیفیت) را نیاز دارند. ضمن این‌که توجه نرمال چندمتغیره، در عمل غالباً مشکل است این نمودارها برای داده‌هایی با ابعاد بالا نیز کاربرد ندارند. در حیطه‌ی  $MSPC$ ، بیشتر تکنیک‌های تحلیل داده‌های چندمتغیره بر اساس فرضیه نرمال چندمتغیره است و روش‌های زیادی برای پرداختن به مسئله غیر نرمال چندمتغیره وجود ندارد. از طرفی، مؤثرترین راهبرد برای مدیران فرایندهای پیچیده با تعداد بُعد

Corresponding author: bamenimoghadam@atu.ac.ir

<sup>1</sup> Multivariate Statistical Process Control

تاریخ دریافت: ۱۴۰۰/۰۹/۰۶ تاریخ پذیرش: ۱۴۰۱/۰۱/۲۷

دوره ۱۲/ شماره ۱

صفحات ۱-۱۴

شامل اندازه‌های کیفیتی چندمتغیره است استفاده کردند و آماره‌هایی بر اساس مفهوم ژرفای داده‌ها بدست آوردند [۱۰]. بی و همکاران (۲۰۱۶) خلاصه‌ای از نتایج پژوهش‌هایی که تا سال ۲۰۱۶ برای ساختن نمودارهای کنترلی چندمتغیره ناپارامتری، که با تمرکز اصلی روی نمودارهای کنترلی داده ژرفاپایه انجام شد، ارائه کردند [۱۱]. چاباک گورسکا (۲۰۱۸) پیش‌نهاد کرد که از نمودارهای کنترلی چندمتغیره‌ای استفاده شود که بر اساس ژرفای داده‌هایی هستند که توسط لیو (۱۹۹۵) پیش‌نهاد شدند و ایده و اصول نمودار کنترلی چندمتغیره بر پایه ژرفای داده‌ها، برای مکان و مقیاس گروه فرعی را معرفی کرد و از آن برای ارزیابی استواری آماری فرایند در یک شرکت تولیدی استفاده کرد [۱۲]. باراله و شیرک (۲۰۱۹) برای کشف یک شیفت در پارامتر مکان فرایند چندمتغیره، با استفاده از آماره‌های آزمون  $T$ -پایه لی و لیو (۲۰۰۴) و  $W_1$ -پایه و شیرک و خورات (۲۰۱۷) که بر اساس ژرفای داده‌ها هستند، نمودارهای کنترلی ناپارامتری را پیش‌نهاد کردند و برای توزیع‌های مختلف، اجرای نمودارهای کنترلی پیش‌نهاد شده را با استفاده از اندازه‌ی میانگین طول اجرا بررسی کردند [۸]. حال با توجه به این که برای پایش پارامتر مکان فرایند چندمتغیره، تعداد معدودی نمودارهای کنترلی ژرفاپایه وجود دارند که با استفاده از روش آزمون فرض مناسب برای داده‌های غیرنرمال پیش‌نهاد شده باشند، و آماره کنترلی برخی از این نمودارها دارای ویژگی‌های مطلوبی مانند ناوردایی آفین و/یا آزادتوزیعی نیستند و برخی نیز حجم محاسباتی بالا دارند، در این مقاله، بر اساس خواص مطلوب توابع ژرفا و با استفاده از ارتباط شناخته شده بین روش‌های نمودار کنترلی و مسئله آزمون فرض، نمودار کنترلی چندمتغیره ناپارامتری نوع شوهراتی پیش‌نهاد می‌شود که خواص مطلوب‌تری دارد. برای ساختن این نمودار کنترلی، از رده‌ی آزمون‌های ژرفاپایه‌ای استفاده می‌شود که کاربرد آن‌ها انجام آزمون نقطه‌ی تقارن (پارامتر مکان) توزیع متقارن چندمتغیره حول  $\mu$ ، در مقابل تقارن حول نقطه‌ی  $\mu_0 \neq \mu$ ، در واقع انجام آزمون فرض

$$\begin{cases} H_0: \mu = \mu_0 \\ H_a: \mu \neq \mu_0 \end{cases}$$

است.

از طرف دیگر، یک بخش با اهمیت و استاندارد در طراحی آماری نمودارهای کنترلی، طراحی نمودار کنترلی با اجرای تحت کنترل ( $IC^2$ ) دلخواه در فاز II است. میانگین طول اجرای تحت کنترل ( $ARL^{(3)}$ )، تعداد متوسط نمونه است قبل از آنکه نمودار کنترلی هشدار یک علت اسنادپذیر در رابطه با کیفیت ضعیف را

در تحلیل‌های آماری چندمتغیره بسیار مفید است. بر اساس این رتبه‌بندی مرکز به بیرون، لیو و همکاران (۱۹۹۹) روش‌های آزمون و تصویری (نموداری) مختلفی را برای تحلیل پارامترهای مختلف توزیع‌های چندمتغیره مانند مکان، مقیاس، چولگی و کشیدگی معرفی کردند [۳]. دو آزمون به نام‌های  $T$ -پایه و  $M$ -پایه برای آزمون کردن مکان‌ها و مقیاس‌های چندمتغیره بر اساس ژرفای داده‌ها، توسط لی و لیو (۲۰۰۴) پیش‌نهاد شدند [۴]. ونچلک (۲۰۱۱) مروری بر کاربردهای ممکن ژرفای داده‌ها که شامل آشکار کردن دورافتاده‌ها، برآوردهای مکان ناوردای آفین و استوار، آزمون‌های رتبه‌ای برای اختلاف مقیاس چندمتغیره، و طبقه‌بندی کننده‌های ژرفاپایه برای حل مسأله‌ی تفکیک را ارائه کرد [۵]. شیرک و خورات (۲۰۱۷) آزمون‌های  $W_1$ -پایه و  $W_2$ -پایه را بر اساس اختلاف‌های ژرفاها معرفی کردند [۶]. دهقان و فریدروحانی (۲۰۱۸) چند خانواده از آزمون‌های ناپارامتری ژرفاپایه برای تقارن مرکزی و نقطه‌ی تقارن (پارامتر مکان) چندمتغیره و آماره‌های آن‌ها که ناوردای آفین و آزادتوزیع هستند را پیش‌نهاد کردند [۷]. برای پایش پارامتر مکان، از بین مراجع ذکر شده نمودارهای کنترلی آزمون‌های  $T$ -پایه،  $W_1$ -پایه و  $W_2$ -پایه توسط باراله و شیرک (۲۰۱۹) معرفی شدند [۸].

روش‌های نمودار کنترلی می‌توانند به‌عنوان مسئله‌ی آزمون فرض مورد مطالعه قرار بگیرند. ناحیه رد در این آزمون‌ها با ناحیه خارج از کنترل در نمودار کنترلی منطبق است. مقادیر نمونه‌ای آماره‌های فرایند تولید با توجه به زمان و مقدار آن‌ها روی یک نمودار رسم می‌شوند. اما تعداد معدودی نمودارهای کنترلی که با استفاده از روش آزمون فرض مناسب برای داده‌های غیر نرمال پیش‌نهاد شده باشند، وجود دارند. مطالعه‌هایی در زمینه‌ی ایجاد نمودارهای کنترلی، بر اساس آزمون فرض‌هایی که به‌طور خاص آماره‌ی آزمون آن‌ها ژرفاپایه هستند، انجام شدند. یک ویژگی مناسب توابع ژرفا، در نظر گرفتن ویژگی‌های هندسی داده‌های چندمتغیره است. از طرفی اکثر توابع ژرفا ناوردای آفین هستند. لیو (۱۹۹۵) اولین نمودارهای کنترلی چندمتغیره‌ی آزادتوزیع، شامل نمودارهای ناپارامتری  $r$  (نمودار نوع شوهراتی برای اندازه‌های منفرد)،  $Q$  (نمودار نوع شوهراتی بر اساس میانگین‌های نمونه‌های فرعی) و  $K$  (نمودار نوع جمع تجمعی) را با استفاده از مفهوم داده‌ژرفا، به منظور پایش فرایند اندازه‌های کیفیت چند متغیره، تعریف کرد [۹]. هامورکارا اوغلو و همکاران (۲۰۰۴) از نمودارهای کنترلی  $Q$  و  $r$  برای کنترل فرایندی که

<sup>3</sup> In-control average run length

<sup>2</sup> In-Control

دریافت کنند. برای مثال، وقتی که یک نمودار کنترلی در حلت کنترل با  $MRL = 256$  طراحی شده باشد، ۵۰ درصد شانس وجود دارد که یک هشدار اشتباه بعد از ۲۵۶ امین نمونه رخ ندهد (و زودتر اتفاق بیفتد). در این مقاله می‌خواهیم به این موضوع بپردازیم که چگونه می‌توان برای کاربری که مایل است فقط ۱۰٪  $P_{100}$  شانس داشتن یک هشدار اشتباه قبل از  $C$  امین نمونه را داشته باشد، نمودار کنترلی طراحی کرد. این نوع طراحی نمودار کنترلی، طراحی مبتنی بر صدک ( $PL^y$ ) نامیده می‌شود. طراحی  $PL$  برای نمودار کنترلی،  $RL$  مشخصی را با احتمال از قبل تعیین شده تضمین می‌کند. به عبارت دیگر، طراحی یک نمودار کنترلی به طوری که صدک  $RL$  داده شده بزرگتر از یک مقدار دلخواه باشد. همچنین طراحی نموداری که کاربر را با طول اجرای خارج از کنترل ( $RL^A$ ) دلخواه، با احتمال از پیش تعیین شده مواجه کند. قابلیت انعطاف‌پذیری روش  $PL$  تضمین می‌کند که کاربران بتوانند نمودارهای کنترلی بر اساس صدک‌های مختلف توزیع طول اجرا طراحی کنند. طراحی بر اساس  $ARL$  را می‌توان به‌عنوان یک مورد خاص از طراحی  $PL$  در نظر گرفت. در این مقاله کارایی طراحی به‌روش  $PL$  با طراحی بر اساس  $ARL$  مقایسه می‌شود. همانطوری که قبلاً اشاره شد، طراحی به‌روش  $PL$  عمومی‌تر بوده و بنابر این نسبت به طراحی‌های  $ARL$  و  $MRL$  انعطاف‌پذیرتر است. همچنین، طراحی  $PL$  تفسیر معنادارتری برای اجراهای تحت کنترل و خارج از کنترل یک نمودار کنترلی می‌دهد. ابزار  $PL$  نمودارهای کنترلی را به طریقی طراحی می‌کند که احتمال یک هشدار بوسیله‌ی تعداد معینی از نمونه‌ها می‌تولند مشخص شود. در مقابل، اندازه‌های  $ARL$  و  $MRL$  فقط به ترتیب میانگین و میانه تعداد نمونه‌ها تا هشدار را فراهم می‌کنند.

در این مقاله با استفاده از کلاس آماره‌های آزمون ژرفاپایه‌ی معرفی شده توسط دهقان و فریدروحانی (۲۰۱۸)، که آن‌ها را به‌عنوان آماره‌های کنترلی در نظر خواهیم گرفت، نتایج ابزار  $PL$  را با ابزار  $ARL$  کلاسیک در طراحی نمودارهای کنترلی مقایسه کرده و در مورد فایده‌ی مفهومی ابزار  $PL$  که صدک‌های طول اجرا با احتمال از قبل تعیین شده تضمین می‌شوند، بحث می‌کنیم. آماره‌های کنترلی ژرفاپایه‌ی مورد نظر دارای خاصیت ناوردایی آفین و آزاد توزیع هستند. ویژگی ناوردایی آفین تضمین می‌کند که عملکرد نمودار کنترلی به سیستم مختصات پایه و

بدهد و اجرای تحت کنترل اغلب توسط آن اندازه‌گیری می‌شود. زمانی که فرایند خارج از کنترل باشد، میانگین طول اجرای خارج از کنترل ( $ARL^4$ ) تعداد متوسط مشاهدات متوالی، تا صدور هشدار درست است. به‌طور معمول، در طراحی به دنبال نمودار کنترلی هستیم که یک علت اسنادپذیر مربوط به کیفیت ضعیف را با طول اجرای خارج از کنترل کوچک هشدار دهد که متضمن تحمیل یکسری شرایط بر میانگین طول اجرای خارج از کنترل است. این روش برای نمودار کنترلی شوهار تی  $\bar{X}$  و  $S^2$  که شاید امروزه به دلیل قابلیت استفاده و سودمندی آنها، تقریباً معمول‌ترین نمودارهایی هستند که برای کنترل فرایند آماری استفاده می‌شوند، در مقاله فراز و همکاران با توضیحات کامل نشان داده شده است [۱۳]. اگر نمونه‌های متوالی مستقل باشند، توزیع طول اجرا هندسی است. زمانی که فرایند تحت کنترل باشد، طول اجرا تعداد نمونه‌ها قبل از هشدار اشتباه است و احتمال آن که یک نمونه‌ی انفرادی خارج از کنترل رسم شود دقیقاً برابر  $\alpha$  (نرخ خطای نوع اول) است. از آن‌جا که توزیع هندسی با تغییرپذیری استاندارد بزرگی، چوله به راست است، براحتی می‌توان متوجه شد که ممکن است  $ARL$  یک اندازه مناسب برای اجرای نمودار کنترلی نباشد، به این معنا که با احتمال ضعیفی امکان دارد طول اجرای تحت کنترل واقعی ( $RL^5$ ) بطور غیر قابل قبولی کوچک باشد. در واقع این نکته توسط مونتگومری [۱۴]، که به استفاده از  $ARL$  به‌عنوان یک اندازه‌ی اجرا برای طراحی نمودارهای کنترلی انتقاد می‌کند، بررسی شده است. مونتگومری [۱۴] به دو دغدغه بزرگ در استفاده از  $ARL$  به‌عنوان یک اندازه‌ی اجرا اشاره می‌کند. اول، آنکه توزیع طول اجرا هندسی، با انحراف استاندارد بزرگ است و دوم آن که توزیع طول اجرا به‌شدت چوله است. از آن‌جا که امکان دارد تفسیرها بر اساس اندازه  $ARL$  گمراه‌کننده باشند، بعضی از پژوهشگران از میانه طول اجرا ( $MRL^6$ ) برای اندازه‌گیری اجرای نمودار کنترلی استفاده کرده‌اند [۱۵]، [۱۶]، [۱۷]، [۱۸]، [۱۹]، [۲۰] و [۲۱]. همه آن‌ها به این نتیجه رسیده‌اند که  $MRL$  اندازه مناسب‌تری برای اجرای نمودار است زیرا کمتر تحت تأثیر چولگی توزیع طول اجرا است. با این وجود با توجه به آن که  $MRL$ ، ۵۰ امین صدک توزیع طول اجرا است، امکان دارد ۵۰ درصد از اوقات (که نسبت قابل ملاحظه‌ای از زمان است)، کاربران هشدارهای اشتباه را زودتر از زمانی که طراحی کرده‌اند،

<sup>7</sup> Percentile-based design

<sup>8</sup> Out-of-control run length

<sup>4</sup> Out-of-control average run length

<sup>5</sup> In-control run length

<sup>6</sup> Median run length

(P۲) بیشترین بودن در مرکز: اگر توزیع  $G$  نسبت به نقطه  $\theta$  متقارن باشد (منظور یکی از مفاهیم تقارن چندمتغیره است) در این صورت  $D(x, G)$  در  $\theta$  ماکزیمم می شود،

$$D(\theta, G) = \sup_{x \in \mathbb{R}^p} D(x, G)$$

(P۳) یکنواخت بودن نسبت به ژرف ترین نقطه: اگر برای هر  $x \in \mathbb{R}^p$ ،  $D(\theta, G) \geq D(x, G)$

آن گاه برای هر  $x \in \mathbb{R}^p$  و  $\alpha \in [0, 1]$

$$D(\theta + \alpha(x - \theta), G) \geq D(x, G)$$

(P۴) به صفر رسیدن در بینهایت: اگر  $\|x\| \rightarrow \infty$ ، آن گاه

$$D(x, G) \rightarrow 0.$$

تعریف: ژرفای ماهالانوبیس ( $MD^{11}$ ) نقطه‌ی  $x \in \mathbb{R}^p$  نسبت به توزیع  $G$  به صورت

$$MD(x, G) = \frac{1}{1 + (x - \mu)^T \Sigma^{-1} (x - \mu)}$$

تعریف می شود که  $\mu$  و  $\Sigma$  به ترتیب، بردار میانگین و ماتریس پراکندگی توزیع  $G$  هستند. نسخه‌ی نمونه‌ای ژرفای ماهالانوبیس با جایگزین کردن  $\mu$  و  $\Sigma$  با برآوردهای نمونه‌ای آن‌ها به دست می آید [۲۳]. زو و سرفلینگ (۲۰۰۰) نشان دادند که چنانچه  $\mu$  نقطه‌ی تقارن تابع توزیع  $G$  باشد، تابع ژرفای ماهالانوبیس در هر چهار ویژگی P۱ تا P۴ صدق می کند [۲۲].

## ۲-۲- آماره کنترلی ژرفاپایه

دهقان و فریدروحانی (۲۰۱۸) متناظر با نرم اقلیدسی مشاهدات که اندازه‌ای از دورافتادگی نقاط است، تابع دورافتادگی نمونه‌ای در نقطه‌ی  $x \in \mathbb{R}^p$  را به نحوی که مقدار آن در بازه‌ی  $[0, \infty]$  باشد، به صورت

$$O(x, \hat{G}_n^s) = \frac{1}{D(x, \hat{G}_n^s)} - \frac{1}{a}$$

توزیع تجربی نمونه‌ی متقارن شده‌ی  $\pm X_1, \dots, \pm X_n$  است و  $\hat{G}_n^s \in [0, a]$  و  $a \in (0, 1]$  در نظر گرفتند که تابع  $O(x, \hat{G}_n^s)$  تعریف کردند

$$Q_{ni} = \frac{X_i}{\sqrt{1 + O(X_i, \hat{G}_n^s)}}.$$

رده‌ی آماره‌های آزمون پیشنهادی را در قالب صورت درجه دوم بر اساس  $Q_{ni}$  ها  $i = 1, \dots, n$  به صورت

$$T_n = n \bar{Q}_n^T (n^{-1} \sum_{i=1}^n Q_{ni} Q_{ni}^T)^{-1} \bar{Q}_n \quad (1)$$

مقیاس‌های اندازه‌گیری بستگی ندارد. همچنین، فایده ابزار PL بر حسب داشتن زمان تا هشدار مشابه در حالت خارج از کنترل ( $TS_1^1$ ) در نقاط درصدی مختلف توزیع زمان تا هشدار، برای مقادیر زمان تا هشدار بهتر در حالت تحت کنترل ( $TS_1^1$ ) را نشان می‌دهیم. در ضمن فرض بر آن است که طراحی در فاز II، یعنی وضعیتی که پارامترهای یک فرایند ایستا شناخته شده هستند، رخ می‌دهد. در واقع در فاز I مطالعه ارزیابی اجرای فرایند و تحلیل قابلیت فرایند انجام شده است. تحلیل‌های بعدی بر نوع و اندازه (بزرگی) دلایل اسنادپذیر که می‌توانند رخ دهند دلالت می‌کنند.

## ۲- معرفی آماره کنترلی ژرفاپایه

### ۲-۱- تابع ژرفا

فرض کنید  $X = (X_1, \dots, X_p)^T$  یک بردار تصادفی روی فضای احتمال  $(\Omega, \mathcal{F}, P)$  و  $G$  تابع توزیع متناظر با اندازه احتمال  $P$  بر فضای  $\mathbb{R}^p$  باشد. بر پایه تابع توزیع  $G$  روی  $\mathbb{R}^p$ ، تابع ژرفا به صورتی که ترتیبی (رتبه‌بندی) مرکز به بیرون از نقاط  $X$  متعلق به  $\mathbb{R}^p$  نسبت به توزیع  $G$  را فراهم کند، تعریف می‌شود. مرکزیت یا دورافتادگی مشاهدات داده شده با توجه به تابع توزیع  $G(\cdot)$  یا ابر داده‌های  $X = (X_1, \dots, X_n)$  می‌تواند با استفاده از نظریه ژرفای داده‌ها اندازه‌گیری شود. ژرفای نقطه‌ی  $x \in \mathbb{R}^p$  میزان عمق/ مرکزیت نقطه‌ی  $x$  را با در نظر گرفتن تابع توزیع  $G(\cdot)$  یا ابر داده‌های  $X$  اندازه می‌گیرد. هر چه مقدار ژرفا بزرگ‌تر باشد، عمق مشاهده متناظر، با توجه به توزیع  $G(\cdot)$  یا ابر داده‌های  $X$  بیش‌تر است. بنابراین یک ترتیبی مرکز به بیرون از نقاط فراهم شده است که می‌تواند برای تحلیل‌های آماری چندمتغیره مفید باشد. ژرفای هر نقطه‌ی  $x \in \mathbb{R}^p$  با در نظر گرفتن توزیع  $G(\cdot)$  یا ابر داده‌های  $X$  توسط تابع ژرفای مناسب  $D(\cdot, G)$  محاسبه شده است. زو و سرفلینگ (۲۰۰۰) تابع ژرفای آماری را به صورت نگاشت نامنفی، کران‌دار که چهار خاصیت مطلوب را تأمین می‌کند، تعریف می‌کنند [۲۲]. این خواص عبارت‌اند از:

(P۱) **ناوردایی آفین**: برای هر ماتریس  $p \times p$  نامنفرد و هر بردار  $p$  بعدی  $b$ .

$$D(Ax + b, G_{Ax+b}) = D(x, G)$$

که  $G_{Ax+b}$  تابع توزیع بردار تصادفی  $AX + b$  است.

<sup>11</sup> Mahalanobis Depth

<sup>10</sup> Out-of-control time to signal

<sup>10</sup> In-control time to signal

که

$$\bar{Q}_n = \frac{1}{n} \sum_{i=1}^n Q_{ni}$$

دهقان و فریدروحانی (۲۰۱۸) نشان دادند که طول بردارهای ژرفاپایه پیشنهادی  $Q_{ni}$  به میزان دوری نقاط اولیه از مبدأ بستگی دارد، و این بردارها اطلاعات مربوط به بزرگی (فاصله‌های استوار از مرکز) و جهت (علامت با توجه به مرکز) نقاط اولیه را لحاظ می‌کنند. این ویژگی  $Q_{ni}$  تضمین می‌کند که ساختار هندسی مجموعه‌ی داده‌های اولیه حفظ می‌شود. علاوه براین،  $Q_{ni}$ ها نسبت به داده‌های اولیه متمرکزتر هستند، که در مواردی مانند توزیع‌های دم‌سنگین که تحت آن‌ها داده‌های بسیار پراکنده‌ای تولید می‌شوند، انتظار می‌رود افزایش یابد. خاصیت‌های ناوردایی آفین و آزاد توزیع بودن  $T_n$  نیز دو ویژگی مطلوب این آماره هستند [۱۷].

### ۲-۲-۱- توزیع مجانبی تحت فرض صفر

در این بخش توزیع مجانبی  $T_n$  تحت فرض صفر برای رده‌ی توزیع‌های متقارن مرکزی معرفی می‌شود. ابتدا بردار تصادفی

$$V_i = \frac{X_i}{\sqrt{1 + O(X_i, G)}}$$

تعریف شده است که  $V_i$  با جایگذاری  $G$  با  $G_n^s$  در  $Q_{ni}$  حاصل شده است. از لم زیر استفاده می‌شود.

لم ۱: فرض کنید  $X_1, \dots, X_n$  نمونه‌ای تصادفی از توزیع متقارن مرکزی حول صفر باشد و گشتاورهای دوم مؤلفه‌های  $X_i$  وجود داشته و متناهی باشد. اگر تابع ژرفای نمونه‌ای در شرط سازگاری یکنواخت

$$\lim_{n \rightarrow \infty} \text{Sup} |D(x, \hat{G}_n) - D(x, G)| = 0, x \in R^p$$

صدق کند و ویژگی  $(P_1)$  برای تابع ژرفا برقرار باشد، آن‌گاه

$$\sqrt{n}(\bar{Q}_n - \bar{V}_n) = O_p \tag{۱}$$

$$\bar{V}_n = \frac{1}{n} \sum_{i=1}^n V_i$$

که

و داریم

$$\frac{1}{n} \sum_{i=1}^n Q_{ni} Q_{ni}^T - \frac{1}{n} \sum_{i=1}^n V_i V_i^T = O_p(1)$$

همچنین شرط سازگاری یکنواخت برای تابع ژرفای ماهالانویس تحت متناهی بودن گشتاورهای مرتبه دوم برقرار است [۲۳].

قضیه ۱: فرض کنید شرایط لم ۱ برقرار باشد، در این صورت تحت فرض  $H, T_n$  دارای توزیع مجانبی  $\chi^2_{p, \alpha}$  درجه آزادی است.

$$T_n \xrightarrow{d} \chi^2_p$$

این قضیه نشان می‌دهد که هرگاه آماره‌ی  $T_n$  از صدک  $100\alpha$ م بالایی توزیع  $\chi^2_{p, \alpha}$  آزادی، تجاوز کند فرض صفر رد خواهد شد.

### ۳- طراحی نمودار کنترلی $T_n$ با ابزار مبتنی بر صدک

فرایندی با دو مشخصه‌ی کیفیت که دارای توزیع نرمال با میانگین تحت کنترل  $\mu$  و ماتریس کوواریانس  $\Sigma$  باشند، را در نظر بگیرید. در این موارد، ابزار معمول در طراحی نمودارهای کنترلی شوهارتی این است که ابتدا نرخ خطای نوع اول در سطح  $\alpha$  را در نظر گرفته و سپس حدود کنترلی را محاسبه می‌کنیم. برای نمودار کنترلی نوع شوهارتی معمول، مانند نمودار کنترلی  $T^2$ -هتلینگ، از حدود کنترلی ثابت که بر اساس توزیع آماره‌های کنترلی هستند استفاده می‌شود. شبیه به نمودارهای کنترلی چندمتغیره مختلف، نمودار کنترلی  $T_n$  فقط یک حد کنترلی دارد. از آن‌جا که آماره  $T_n$  تحت فرض  $H$  دارای توزیع مجانبی  $\chi^2_{p, \alpha}$  صدک  $100\alpha$ م بالایی توزیع  $\chi^2_{p, \alpha}$  آزادی است و چون آماره آزمون  $T_n$  شاخص تعمیم یافته‌ای از فاصله است، لذا حد پایین نمودار کنترلی مورد نظر همواره صفر است.

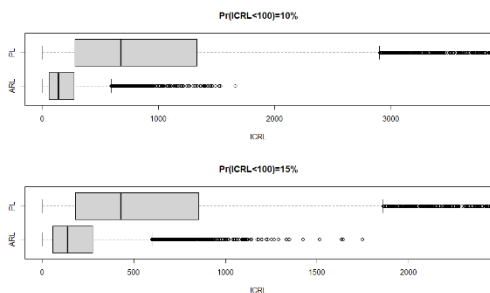
پس

$$LCL = 0 \text{ و } UCL = \chi^2_{\alpha}(p) \tag{۲}$$

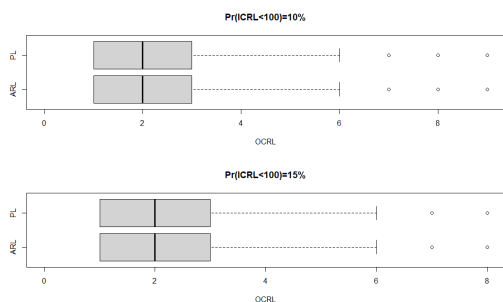
به این ترتیب، نمودار کنترلی پیشنهادی یک وضعیت خارج از کنترل را نشان می‌دهد اگر مقدار آماره‌ی  $T_n$  بزرگ‌تر یا مساوی  $UCL$  باشد. در طراحی نمودارهای کنترلی مبتنی بر صدک به جای اندازه  $ARL$ ، نمودارهای کنترلی بر اساس صدک دلخواه توزیع  $RL_{P_1}$  طراحی می‌شوند [۱۳]. برای این کار، ابتدا کاربر احتیاج دارد در مورد اجرای تحت کنترل دلخواه با دو پارامتر  $C$  (مقدار دلخواه برای  $RL$ ) و سطح احتمال  $P_1$  تصمیم بگیرد. بنابراین ثابت‌های نمودار کنترلی باید طوری تنظیم شوند که  $P_1$  امین صدک  $RL_{P_1}$  مساوی یا بزرگ‌تر از مقدار دلخواه  $C$  باشد، به این معنی که،  $Pr(RL_{P_1} < C) = P_1$  یا  $RL_{P_1} = C$ . از آن‌جا که توزیع  $RL$  هندسی است، نشان دادن این که  $P_1$  امین صدک  $RL$  می‌تواند به وسیله‌ی  $RL_{P_1} = \frac{\ln(1-P_1)}{\ln(1-\alpha)}$  محاسبه

باشیم که بهبود عملکرد تحت کنترل، به طراحی نمودار کنترلی با رویکرد PL ارتباط دارد و به از بین رفتن قابلیت تشخیص نمودار کنترلی مرتبط نیست. یعنی در رویکرد PL مقدار متوسط طول اجرای تحت کنترل ARL تقریباً ۴/۷ برابر رویکرد ARL است و این افزایش مقدار ARL به معنی از دست رفتن قابلیت تشخیص نمودار کنترلی نیست.

همان‌طور که این مثال نشان می‌دهد، در طراحی آماری نمودارهای کنترلی، یک مزیت آشکار از رویکرد PL نسبت به رویکرد ARL وجود دارد. نتایج گویای این مسئله است که نمودار کنترلی مبتنی بر ابزار PL کارایی تحت کنترل بهتری را نسبت به نمودار مبتنی بر ابزار ARL دارد و در صورت اتفاق افتادن تغییر در میانگین فرایند، توانایی کشف این تغییر را به اندازه‌ی ابزار ARL دارد. در واقع در مقابل به وجود آمدن تغییر در میانگین فرایند، اجرایی به‌خوبی نمودار مبتنی بر ARL، همراه با اجرای تحت کنترل خیلی بهتر، فراهم می‌کند.



شکل ۱- نمودار جعبه‌ای ARL برای نمودار کنترلی  $T_n$  با رویکردهای PL و ARL برای  $n = 50$  و مقادیر مختلف سطح احتمال  $P_1$ .



شکل ۲- نمودار جعبه‌ای ARL برای نمودار کنترلی  $T_n$  با رویکردهای PL و ARL برای  $n = 50$  و  $d = 3$  و مقادیر مختلف سطح احتمال  $P_1$ .

شود، آسان است. پس می‌توان نرخ خطای نوع I دلخواه را طوری در نظر گرفت که  $RL_{P_1} = C$  به عبارت دیگر،

$$\alpha = 1 - \exp\left(\frac{\ln(1-P_1)}{C}\right) \quad (3)$$

برای مثال، فرض کنید که یک فرایند با دو مشخصه ( $p = 2$ ) دارای توزیع نرمال با پارامترهای  $\mu = 0$  و  $\Sigma = I$  باشد. ممکن است بخواهیم نمودارهای کنترلی شوهراتی با حجم نمونه‌های  $n = 5$  مستقل از هم طوری طراحی کنیم که ۸۰ درصد از اوقات  $RL > 120$  یا  $\Pr(RL > 120) = 0.80$  بنابراین، نمودار کنترلی باید طراحی شود که صدک ۲۰ ام  $RL_{.20}$  حداقل ۱۲۰ باشد. در این حالت، باید نمودارهای کنترلی با مقدار خطای نوع I داده شده در (۳) ساخته شوند،

$$\alpha = 1 - \exp\left(\frac{\ln(1-0.20)}{120}\right) = 0.00185$$

یعنی بالای نمودار کنترلی  $T_n$  شوهراتی به صورت  $K = \chi^2_{\alpha}(p) = 13.815$  بدست می‌آید. نرخ خطای نوع II برای نمودار کنترلی مورد نظر، به منظور کشف شیفت و انتقال در میانگین فرایند به اندازه‌ی  $d$ ، از طریق شبیه‌سازی مونت کارلو بدست می‌آید. شکل ۱ اجرا و کارایی تحت کنترل نمودار کنترلی  $T_n$  مبتنی بر رویکرد ARL برای  $n = 50$  و  $\alpha = 0.005$  را با نمودار کنترلی مبتنی بر رویکرد PL با  $n = 50$  و  $\alpha$ ی رابطه‌ی (۳)، در دو حالت:  $\Pr(RL > 100) = 0.90$ ،  $\Pr(RL > 100) = 0.85$  مقایسه می‌کند (از آن‌جا که توزیع مجانبی آماره‌ی کنترلی موردنظر تحت فرض صفر خنثی دو است، اجرای نمودار کنترلی  $T_n$  برای  $n \geq 25$  انجام شده است). به‌عنوان مثال، برای حالت اول ( $P_1 = 0.1$ )، میزان  $\alpha$  با استفاده از رابطه‌ی (۳) برابر

$$\alpha = 1 - \exp\left(\frac{\ln(0.1/90)}{100}\right) = 0.00105$$

صورت ضریب حدود کنترلی نمودار  $T_n$  با رویکرد PL برابر  $K = \chi^2_{0.00105}(2) = 13.71793$  محاسبه می‌شود. مقادیر متوسط طول اجرای تحت کنترل و میانه طول اجرای تحت کنترل با رویکرد ARL ( $\alpha = 0.005$ )، به ترتیب برابر

$$MRL = \frac{\log(0.5)}{\log(1-\alpha)} = 128/28 \quad \text{و} \quad ARL = \frac{1}{\alpha} = 200$$

است. این مقادیر با رویکرد PL به ترتیب برابر

$$MRL = \frac{\log(0.5)}{\log(1-\alpha)} = 659/79 \quad \text{و} \quad ARL = \frac{1}{\alpha} = 952/38$$

است. در نتیجه رویکرد PL کارایی تحت کنترل را بهبود می‌بخشد و با این رویکرد می‌توان پیشرفت بسیار زیادی در عملکرد تحت کنترل نمودارهای کلاسیک به‌دست آورد. شکل ۲ عملکرد خارج از کنترل نمودار کنترلی  $T_n$  را برای آشکار کردن انتقال در میانگین به اندازه  $d = 3$  با دو رویکرد ARL و PL و مقادیر مختلف سطح احتمال مقایسه می‌کند. باید توجه داشته

#### ۴- طراحی آماری نمودار کنترلی $T_n$ با ابزار مبتنی بر صدک

در عمل ترجیح می‌دهیم طراحی آماری PL ی را فراهم کنیم که با احتمالات معینی  $RL_1$  و  $RL_2$  تضمین شوند [۱۳]. در این حالت، پارامتر  $h$  که فاصله‌ی نمونه‌گیری است را معرفی کرده و با استفاده از  $h$ ، طول اجرا را به واحدهایی از زمان تبدیل می‌کنیم. به عبارت دیگر، پارامترهای  $n$ ،  $h$  و  $k$  را طوری پیدا می‌کنیم که زمان تا هشدار تحت کنترل و خارج از کنترل با احتمال‌های معینی تضمین و به صورت زیر محاسبه شوند:

$$TS_1 = RL_1 * h; RL_1 \sim Ge(\alpha)$$

$$TS_2 = RL_2 * h; RL_2 \sim Ge(1 - \beta)$$

که  $Ge(r)$  نشانگر توزیع هندسی با پارامتر  $r$  و  $TS_1$  و  $TS_2$  به ترتیب زمان تا هشدار تحت کنترل و زمان تا هشدار خارج از کنترل هستند. طراحی آماری به‌صورتی که در ادامه آمده است، تعریف می‌شود. تابع هدف کمینه کردن صدک  $100P_2$  ام بالای زمان تا هشدار خارج از کنترل  $TS_{1P_2}$  است به طوری که اجرای دلخواه حاصل شود. متغیرهای تصمیم  $\theta = \{n, h, k\}$  و  $T_1$  و  $T_2$  به ترتیب مقادیر دلخواه  $TS_1$  و  $TS_2$  هستند.

$$\text{Min } TS_{1P_2}; \theta = \{k, n; h\}$$

در رابطه با:

$$Pr(TS_1 \geq T_1) = 1 - P_1$$

$$Pr(TS_2 \leq T_2) = 1 - P_2$$

$$k > 0; TS_1 > TS_2$$

$$n \in \{2, 3, \dots, n_{max}\}$$

$$h \in \{h_{min}, 2h_{min}, 3h_{min}, \dots, T_2\}$$

برای آن که طراحی عملی شود، شرط  $h < T_2$  را هم روی فاصله نمونه‌گیری  $h$ ، قرار می‌دهیم.

در این مقاله، از ابزار الگوریتم ژنتیک ( $GA^{12}$ ) که به‌طور موفقیت‌آمیزی در حل مسائل بهینه‌سازی پیچیده بکار گرفته می‌شود، برای حل مسئله بهینه‌سازی (۴) استفاده می‌کنیم و برای یافتن حل بهینه، الگوریتم ژنتیک را در نرم افزار R با استفاده از تابع  $ga()$  کدگذاری می‌کنیم. برای اطلاع بیشتر در مورد الگوریتم ژنتیک (GA) به [۲۴] رجوع شود. برای حل مسئله بهینه‌سازی (۴) الگوریتم ژنتیک به‌صورت زیر است:

۱- تنظیم مقادیر پارامترهای اجرا و فرایند  $T_1, T_2, P_1, P_2, d$ .

۲- تعیین کران‌های بالا و پایین برای حجم نمونه و پارامترهای نمونه‌گیری. کران‌ها باید از لحاظ طراحی، امکان‌پذیر باشند که نمونه‌گیری به تعداد  $n_{max}$  نمونه در هر  $h_{min}$  واحد زمان انجام

شود. اگر طراحی شدنی و امکان‌پذیر نباشد، کران‌ها دوباره تعیین شوند.

۳- تولید ۱۰۰ کروموزوم تصادفی  $(n, h)$  به عنوان جامعه اولیه. هر کروموزوم در واقع یک حل امکان‌پذیر برای طراحی آماری است.

۴- در نظر گرفتن مقدار مورد نظر برای طول اجرای تحت کنترل از رابطه  $C = T_1/h$ .

۵- با داشتن  $C$ ، بدست آوردن  $\alpha$  به کمک معادله‌ی (۳)، و سپس قرار دادن  $k = \chi^2_{\alpha}(p)$ . که  $\chi^2_{\alpha}(p)$  صدک  $100\alpha$  ام بالایی توزیع  $\chi^2$  دو با  $p$  درجه آزادی است. به این ترتیب طول اجرای تحت کنترل دلخواه برای هر نمودار با احتمال  $(1 - P_1)$  تضمین می‌شود.

۶- محاسبه‌ی خطای نوع II  $(\beta_{T_n})$  برای هر کروموزوم با استفاده از شبیه سازی مونت کارلو.

۷- محاسبه‌ی  $(1 - P_2)$  امین صدک زمان تا هشدار در حالت خارج از کنترل توسط  $TS_{1P_2} = \frac{\ln P_2}{\ln \beta} \times h$  برای هر کروموزوم تا تابع برازندگی تعریف شده در (۴) بدست آید. در رابطه مورد نظر  $\beta = \beta_{T_n}$  است.

۸- تعیین عدد بزرگی به‌عنوان جریمه برای مقدار برازندگی مربوط به کروموزوم‌هایی که  $TS_{1P_2} > T_2$ .

۹- مرتب کردن کروموزوم‌ها بر اساس مقدار برازندگی‌شان  $(TS_{1P_2})$ . محاسبه مقدار احتمال انتخاب برای هر کروموزوم به نسبت مقدار برازندگی آنها، که شانس بیشتری را به کروموزوم‌هایی با مقدار برازندگی بهتر می‌دهد.

۱۰- انتخاب کروموزوم برتر در تولید جاری برای آن که حفاظی بر بهترین پاسخ داشته باشیم و کپی کردن آن به‌طور مستقیم در تولید بعدی.

۱۱- انتخاب دو کروموزوم با استفاده از گردونه شانس به عنوان والدین برای ایجاد دو کروموزوم جدید به عنوان فرزندان برای تولید بعدی و تکرار این مرحله تا تولد ۹۰ کروموزوم جدید. فرزندان با عملگرهای متقاطع و جهش ایجاد می‌شوند که میزان متقاطع و جهش در این جا ۵۰ درصد تعیین شده است.

۱۲- تکرار مراحل ۵ تا ۱۰ تا زمانی که دیگر هیچ بهبودی در مقدار برازندگی بهترین کروموزوم در ۱۰۰۰ نسل متوالی وجود نداشته باشد.

<sup>12</sup> Genetic Algorithm

## ۴-۱- تحلیل عددی

این فرضیات بدون از دست دادن هیچ کلیتی در نظر گرفته شده‌اند. جدول‌های ۱ تا ۴ طراحی آماری بهینه برای نمودار  $T_n$  شوهارتی با رویکرد PL با زمان تا هشدار تحت کنترل دلخواه  $T_1$  و زمان تا هشدار خارج از کنترل  $T_2$  برای ترکیبات مختلف پارامترهای فرایند را نشان می‌دهند. نتایج برای موارد  $P_1 = P_2 = 0.05, 0.1, 0.15, 0.2, n_{max} = 50$  و  $T_1 = 100$  واحد از زمان  $T_2, 1, 1/5, \dots, 0.5$  و  $T_2 = 1, 2, 3, 4, 5$  واحد از زمان، برای سه انتقال مختلف ( $d = 1, 2, 3$ ) در میانگین توضیح داده شده است. نتایج بدست آمده در جدول‌های ۱ تا ۴، مقدارهای بهینه پارامترهای تنظیمی نمودار کنترلی  $T_n$ ، یعنی ضریب حدود کنترلی ( $k$ )، اندازه‌ی نمونه ( $n$ ) و فاصله‌ی نمونه‌گیری ( $h$ ) را نشان می‌دهند.

به‌عنوان یک مورد عملی و کاربردی، فاصله‌های نمونه‌گیری ممکن را به‌صورت واحدهای زمان  $h = 0.5, 1, 1.5, \dots, T_2$  تنظیم می‌کنیم. یعنی، حداکثر فاصله نمونه‌گیری روی زمان تا هشدار خارج از کنترل دلخواه در یک شیفت کاری و کوتاه‌ترین فاصله‌ی نمونه‌گیری ممکن  $h_{min} = 0.5$  در هر  $0.5$  واحد از زمان در نظر گرفته شده است. به‌علاوه، فرض می‌کنیم حجم نمونه‌ی ممکن در هر  $0.5$  واحد زمان تا  $50$  ( $n_{max} = 50$ ) و تولید، به ازای این مقدار حداقل برابر  $T_2$  باشد. به عبارت دیگر، حداکثر نرخ نمونه‌گیری  $50$  مورد در هر  $0.5$  واحد از زمان و حداقل نرخ نمونه‌گیری  $35$  مورد در هر  $T_2$  واحد از زمان باشد.

جدول ۱: طراحی آماری بهینه برای نمودار کنترلی  $T_n$  با رویکرد PL،  $n_{max} = 50$  و  $Pr(TS_1 > 100) = Pr(TS_1 \leq T_2) = 0.05$

$d$	۱			۲			۳		
	$k$	$n$	$h$	$k$	$n$	$h$	$k$	$n$	$h$
$T_2$									
۱	۱۵/۱۵	۵۲	۱/۰	۱۵/۱۵	۵۲	۱/۰	۱۵/۱۵	۵۳	۱/۰
۲	۱۴/۳۴	۵۱	۱/۵	۱۴/۳۴	۵۲	۱/۵	۱۵/۱۵	۵۲	۱/۰
۳	۱۴/۳۴	۴۸	۱/۵	۱۴/۳۴	۵۴	۱/۵	۱۴/۳۴	۵۳	۱/۵
۴	۱۴/۳۴	۴۰	۱/۰	۱۴/۳۴	۵۵	۱/۵	۱۴/۳۴	۴۹	۱/۵
۵	۱۵/۱۵	۴۸	۱/۵	۱۳/۷۷	۴۹	۲/۰	۱۴/۳۴	۵۱	۱/۵

جدول ۲: طراحی آماری بهینه برای نمودار کنترلی  $T_n$  با رویکرد PL،  $n_{max} = 50$  و  $Pr(TS_1 > 100) = Pr(TS_1 \leq T_2) = 0.10$

$d$	۱			۲			۳		
	$k$	$n$	$h$	$k$	$n$	$h$	$k$	$n$	$h$
$T_2$									
۱	۱۳/۷۱	۴۷	۱/۰	۱۳/۷۱	۵۱	۱/۰	۱۳/۷۱	۵۵	۱/۰
۲	۱۲/۹۰	۴۹	۱/۵	۱۲/۹۰	۵۳	۱/۵	۱۲/۹۰	۵۵	۱/۵
۳	۱۲/۳۳	۴۹	۲/۰	۱۳/۷۱	۵۱	۱/۰	۱۲/۹۰	۵۳	۱/۵
۴	۱۳/۷۱	۴۸	۱/۰	۱۲/۹۰	۵۴	۱/۵	۱۳/۷۱	۵۴	۱/۰
۵	۱۳/۷۱	۴۸	۱/۰	۱۲/۳۳	۵۳	۲/۰	۱۲/۹۰	۵۵	۱/۵



جدول ۳: طراحی آماری بهینه برای نمودار کنترلی  $T_n$  با رویکرد PL،  $n_{max} = 50$  و  $\Pr(TS_1 \leq T_r) = 0.85$  و  $\Pr(TS > 100) = 0.05$

$d$	۱			۲			۳			
	$T_r$	$k$	$n$	$h$	$k$	$n$	$h$	$k$	$n$	$h$
۱	۱۲/۸۵	۵۵	۱/۰	۱۲/۸۵	۵۴	۱/۰	۱۲/۸۵	۵۴	۱/۰	۱/۰
۲	۱۲/۸۵	۴۸	۱/۰	۱۲/۸۵	۵۰	۱/۰	۱۲/۸۵	۵۳	۱/۰	۱/۰
۳	۱۲/۰۴	۴۹	۱/۵	۱۲/۰۴	۵۳	۱/۵	۱۲/۸۵	۵۴	۱/۰	۱/۰
۴	۱۲/۸۵	۴۸	۱/۰	۱۲/۰۴	۵۰	۱/۵	۱۲/۰۴	۵۴	۱/۵	۱/۵
۵	۱۲/۰۴	۵۱	۱/۵	۱۲/۸۵	۵۲	۱/۰	۱۲/۰۴	۵۲	۱/۵	۱/۵

جدول ۴: طراحی آماری بهینه برای نمودار کنترلی  $T_n$  با رویکرد PL،  $n_{max} = 50$  و  $\Pr(TS_1 \leq T_r) = 0.80$  و  $\Pr(TS > 100) = 0.05$

$d$	۱			۲			۳			
	$T_r$	$k$	$n$	$h$	$k$	$n$	$h$	$k$	$n$	$h$
۱	۱۲/۲۱	۵۰	۱/۰	۱۲/۲۱	۵۴	۱/۰	۱۲/۲۱	۵۲	۱/۰	۱/۰
۲	۱۲/۲۱	۴۹	۱/۰	۱۱/۴۰	۵۲	۱/۵	۱۲/۲۱	۵۵	۱/۰	۱/۰
۳	۱۱/۴۰	۴۶	۱/۵	۱۱/۴۰	۵۴	۱/۵	۱۲/۲۱	۵۴	۱/۰	۱/۰
۴	۱۰/۳۸	۴۲	۲/۵	۱۲/۲۱	۵۱	۱/۰	۱۲/۲۱	۵۱	۱/۰	۱/۰
۵	۱۰/۳۸	۴۲	۲/۰	۱۲/۲۱	۵۰	۱/۵	۱۲/۲۱	۵۱	۱/۰	۱/۰

که محدودیت‌های دلخواه در مورد طول اجرای تحت کنترل و طول اجرای خارج از کنترل را با احتمال‌های از پیش تعیین شده برآورد می‌کند. آماره‌ی کنترلی ژرفاپایه‌ی نمودار کنترلی پیشنهاد شده خاصیت ناوردایی آفین دارد. ویژگی ناوردایی آفین تضمین می‌کند که عملکرد نمودار کنترلی به سیستم مختصات پایه و مقیاس‌های اندازه‌گیری بستگی ندارد. نمونه‌هایی از رویکرد مبتنی بر صدک‌های طول اجرا (PL) برای نمودار کنترلی ژرفاپایه‌ی  $T_n$  ارائه شد. با توجه به مقایسه‌ی کارایی نمودار کنترلی  $T_n$  تحت رویکرد مبتنی بر میانگین طول اجرا (ARL) با رویکرد مبتنی بر صدک‌های طول اجرا (PL)، مشخص شد که کارایی نمودار کنترلی  $T_n$  مبتنی بر رویکرد صدک‌های طول اجرا (PL) از کارایی این نمودار کنترلی مبتنی بر رویکرد متوسط طول اجرا (ARL)، بهتر بوده و در بدترین حالت، جایگزینی مناسب برای رویکرد کلاسیک ARL در طراحی آماری نمودارهای کنترلی است. همچنین از آنجا که در این روش، طول اجرا با احتمال‌های از پیش تعیین شده تضمین می‌شود و عملاً از نظر توزیع طول اجرا غالب هستند، همیشه از نظر مفهومی بهبودی از روش کلاسیک ARL محسوب می‌شود. رویکرد مبتنی بر صدک‌های طول اجرا (PL) شکل تعمیم یافته‌ی از رویکردهای مبتنی بر میانه‌ی طول اجرا (MRL) و

بر اساس نتایج به دست آمده، همچنان که سطح اطمینان از ۹۵ درصد به ۸۰ درصد کاهش می‌یابد، پهنای حدود کنترلی نمودار  $T_n$  ( $k$ ) باریک‌تر می‌شود. به علاوه، متوجه شدیم که فاصله‌ی نمونه‌گیری  $h$  با افزایش  $T_r$  افزایش می‌یابد،  $n$  با افزایش  $T_r$  کاهش می‌یابد که همه نتایج مورد انتظار هستند. مطابق با مقادیر  $P_1$  و  $C$  تعیین شده توسط کاربر و مقادیر بهینه در جدول‌ها، حد کنترلی بالای نمودار ( $UCL$ ) مشخص می‌شود. در این نوع طراحی، نمودارهای کنترلی دارای طول اجرای تحت کنترل و خارج از کنترل تضمین شده هستند و نسبت به نمودار کنترلی  $T_n$  که بر اساس متوسط طول اجرا طراحی می‌شود، پایداری بیشتری دارد که همین مسئله باعث عملکرد بهتر نمودار کنترلی می‌شود. همچنین، احتمال ارتکاب هشدار اشتباه ( $\alpha$ ) را به شدت پایین می‌آورد و از آنجا که تولفایی بالاتری را در به نمایش گذاشتن وضعیت واقعی فرایند دارند، اعتماد کاربر به نمودارهای کنترلی بیشتر می‌شود.

## ۵- نتیجه‌گیری

در این مقاله روشی بر اساس رویکرد مبتنی بر صدک‌های طول اجرا (PL)، برای طراحی نمودار کنترلی ژرفاپایه‌ی  $T_n$  ارائه شد. نشریه مهندسی و مدیریت کیفیت

- [12] Czabak-Gorska, I. D. (2018). *Multivariate Control Charts Based on Data Depth for Subgroup Location and Scale*. *CBU International Conference on Innovations in Science and Education*, 1042-1049.
- [13] Faraz, A., Saniga, E., & Montgomery, D. (2019). *Percentile-Based Control Chart Design with an Application to Shewhart  $\bar{X}$  and  $S^2$  Control Charts*. *Quality and Reliability Engineering International*, 35(1), 116-126.
- [14] Montgomery, D. C. (2013). *Introduction to statistical quality control*. 7<sup>th</sup> ed. Hoboken, NJ: John Wiley & Sons.
- [15] Gan, F. F. (2013). *An Optimal Design of EWMA Control Charts Based on Median Run Length*. *Journal of Statistical Computation and Simulation*, 45, 169-184.
- [16] Gan, F. F. (1994). *An Optimal Design of Cumulative Sum Control Chart Based on Median Run Length*. *Communication in Statistics- Simulation and Computation*, 23(2), 485-503.
- [17] Chakraborti, S. (2007). *Run Length Distribution and Percentiles: The Shewhart  $\bar{X}$  Chart with Unknown Parameters*. *Quality Engineering*, 19(2), 119-127.
- [18] Golosnoy, V., & Schmid, W. (2007). *EWMA Control Charts for Monitoring Optimal Portfolio Weights*. *Sequential Analysis (Design Methods and Applications)*, 26(2), 195-224.
- [19] Das, N. (2009). *A Comparison Study of Three Non-parametric Control Charts to Detect Shift in Location Parameters*. *The International Journal of Advanced Manufacturing Technology*, 41(7-8), 799-807.
- [20] Khoo, M. B. C., Wong, V. H., Wu, Z., & Castagliola, P. (2012). *Optimal Design of the Synthetic Chart for the Process Mean Based on Median Run Length*. *IIE Transactions*, 44(9), 765-779.
- [21] Khoo, M. B. C., Wong, V. H., Wu, Z., & Castagliola, P. (2011). *Optimal Design of the Multivariate Synthetic Chart for Monitoring the Process Mean Vector Based on Median Run Length*. *Quality and Reliability Engineering International*, 27(8), 979-1234.
- [22] Zuo, Y., Serfling, R. J. (2000). *General Notions of Statistical Depth Function*. *Annals of Statistics*, 28(2), 461-482.
- [23] Liu, R.Y., Singh, K. (1993). *A Quality Index Based on Data Depth and Multivariate Rank Tests*. *Journal of the American Statistical Association*, 88(421), 252-260.
- [24] Haupt, R. L., Haupt, S. E., Haupt S. E. (1998). *Practical Genetic Algorithms*. 2 New york, Wiley.
- میانگین طول اجرا (ARL) در طراحی آماری نمودار کنترلی است، به این معنا که به صدک ۵۰ام و میانگین طول اجرا محدود نمی‌شود و بنابراین انعطاف‌پذیرتر از رویکردهای MRL و ARL است.

## ۶- مراجع

[1] Hotelling, H. (1947). *Multivariate quality control*. In: *Techniques of statistical analysis*. McGraw-Hill, New York.

[2] Montgomery, D. C. (1996). *Introduction to statistical quality control*. John Wiley & Sons Inc., USA.

[3] Liu, R. Y., Parelius, J., & Singh, K. (1999). *Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference (with discussion)*. *Annals of Statistics*, 27, 783-858.

[4] Li, J., & Liu, R. Y. (2004). *New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth*. *Statistical Science*, 19(4), 686-696.

[5] Vencalek, O. (2011). *Concept of Data Depth and its Applications*. *Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Mathematica*, 50, 111-119.

[6] Shirke, D., & Khorate, S. (2017). *Power Comparison of Data Depth-Based Nonparametric Tests for Testing Equality of Locations*. *Journal of Statistical Computation and Simulation*, 87(8), 1489-1497.

[7] Dehghan, S., & Faridrohani, M. R. (2018). *Affine Invariant Depth-Based Tests for the Multivariate One-Sample Location Problem*. *TEST*, 28(3), 671-693.

[8] Barale, M. S., & Shirke, D. T. (2019). *Nonparametric Control Charts Based on Data Depth for Location Parameter*. *Journal of Statistical Theory and Practice*, 13(3), 41.

[9] Liu, R. Y. (1995). *Control Charts for Multivariate Processes*. *Journal of the American Statistical Association*, 90, 1380-1387.

[10] Hamurkaroglu, C., Mert, M., & Saykan, Y. (2004). *Nonparametric Control Charts Based on Mahalanobis Depth*. *Hacettepe Journal of Mathematics and Statistics*, 33, 57-67.

[11] Bae, S. J., Do, G., & Kvam, P. (2016). *On Data Depth and the Application of Nonparametric Multivariate Statistical Process Control Charts*. *Applied Stochastic Models in Business and Industry*, 32(5), 660-676.

# Statistical Design of the Percentile-Based Depth-Based Multivariate Control Chart

Shadi Nasrollahzadeh

*Department of Statistics, Science and Research branch, Islamic Azad University, Tehran, Iran;*

Mohammad Bameni Moghadam<sup>1</sup>

*Faculty member of Allameh Tabataba'i Department of Statistics, Allameh Tabataba'i University, Tehran, Iran;*

**Abstract:** We introduce a method for the statistical design of a depth-based control chart, using the percentile-based approach. The proposed control chart is affine invariant and is asymptotically distribution-free. Generally, the performance of a control chart is evaluated with the average run length metric. The average run length metric has a geometric distribution skewed to the right with a large standard deviation and may not be a proper measure for evaluating the control chart. Therefore, we use the statistical design method of control charts with the PL approach, which is an improvement and development on classical statistical design. By employing constraints on average run length, the length of in-control and out-of-control performances are guaranteed with predetermined probabilities and we can ensure that the in-control run length exceeds the desired value and the out-of-control run length is less than the desired value. Simulation studies show that the proposed control chart is more efficient than the average run length approach.

**Keywords:** run length, statistical design, statistical process control, percentile-based control chart, multivariate location

## 1- Aim and Introduction

In multivariate statistical process control (MSPC) and multivariate quality control, control charts are a very important tool for monitoring and/or controlling production processes. In this regard, Hotelling's  $T^2$  control charts are widely used to monitor process location vectors in which the multivariate normality assumptions of response variables (quality characteristics) are required [1]. While in practice, multivariate normality justification is often difficult and these charts are inapplicable to high-dimensional data. In the field of MSPC, most multivariate data analysis techniques are based on the multivariate normality assumption, and there are not many methods for the multivariate non-normality problem. On the other hand, the most effective strategy for managers of complex high-dimensional processes that monitor multivariate data is to reduce the dimension of the problem as much as possible, without losing significant information, in relation to the incoming signal. The effective and efficient measurements that provide sufficient data reduction for a large number of variables in a completely nonparametric way are based on data depth. For this reason, recently, data depth-based non-parametric multivariate analysis techniques, which do not require normality and have no constraints on data dimensions, have become more attractive. In this research, focusing on the known nonparametric multivariate control chart techniques, we introduce a method for the statistical design of a depth-based control chart, using the percentile-based approach. Control chart procedures can be studied as the testing of hypothesis problems. Thus a control chart at any point drawn is a representation of an  $\alpha$ -level test of the following hypotheses:  $\begin{cases} H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ H_a: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \end{cases}$  The rejection region in these tests corresponds to the out-of-control region in the control chart. Sample values of statistics of the production process are recorded with respect to time and their values are plotted on a chart. However, it is worth noting that for monitoring the multivariate process location parameter, there are few proposed depth-based control charts using the testing of hypothesis method for non-normal data. In addition, some of their charting statistics have not desirable features such as affine invariance and/or distribution-free properties and some have high computational performances. A suitable feature of depth

<sup>1</sup> Corresponding author: bamenimoghadam@atu.ac.ir

functions is to consider the geometric properties of multivariate data and also, most of the depth functions are affine invariant.

On the other hand, designing a control chart with desired in-control performance is of great importance in phase II and is a standard part of the classical statistical design. The in-control performance is most often measured by the in-control average run length which is the average number of samples before the control chart signals an assignable cause of poor quality even though no cause exists. Given that the successive samples are independent, the run length is geometrically distributed. When the process is in control, the run length is the number of samples before a false signal and the probability of an individual sample plotting out of control is exactly equal to  $\alpha$  or the type I error rate.

It is well known that the geometric distribution is skewed to the right with a large standard variation. Thus, the ARL metric may not be an adequate measure of control chart performance [2]. Since the interpretation based on the ARL measure can be misleading, some authors have used the median run length (MRL) to measure a control chart's performance [3], [4], [5], [6],[7], [8] and [9]. All of them concluded that the MRL is a more suitable measure for a chart's performance because it is less affected by the skewness of the run length distribution. However, the MRL is still the 50<sup>th</sup> percentile of the run length distribution, in the sense that 50% of time, users may still get false alarms sooner than what is planned.

In this paper, we have dealt with the matter that how to design a multivariate depth-based control chart for a user that there is only  $100P_1\%$  chance of having a false alarm before the  $C^{th}$  sample. The control chart designs of this type are called, percentile-based designs (PL). A PL design ensures that the control chart has a particular in-control run length ( $RL_0$ ) that is guaranteed with a predetermined probability. One would also wish to design the chart such that a desired out-of-control run length ( $RL_1$ ) would be met with a predetermined probability. The flexibility of the PL method ensures that users can design control charts based on the different percentiles of the run length distribution. We can consider the ARL-based design as a special case of the PL design, and we show the efficacy of this method when compared to the ARL design. In this paper, by using the PL-based method, we aim to propose a nonparametric multivariate control chart that has more desirable properties.

## 2- Methodology

To construct the proposed PL-based control chart, a class of depth-based test statistics is used, based on the desirable properties of depth functions [10].

We consider the function of outlyingness of  $\mathbf{x} \in \mathbb{R}^p$ , such that is normalized to take values in  $[0, \infty)$  as  $O(\mathbf{x}, \hat{G}_n^s) = \frac{1}{D(\mathbf{x}, \hat{G}_n^s)} - \frac{1}{a}$ , where  $D(\cdot, \hat{G}_n^s) \in [0, a]$  with  $a \in (0, 1]$  and  $\hat{G}_n^s$  is the sample distribution function of  $\pm X_1, \dots, \pm X_n$  and defined  $Q_{ni} = \frac{X_i}{\sqrt{1+O(\mathbf{x}_i, \hat{G}_n^s)}}$ .

Then, we use the test statistic,  $T_n$ , in a quadratic form  $T_n = n\bar{Q}_n^T (n^{-1} \sum_{i=1}^n Q_{ni} Q_{ni}^T)^{-1} \bar{Q}_n$ , where

$$\bar{Q}_n = \frac{1}{n} \sum_{i=1}^n Q_{ni}.$$

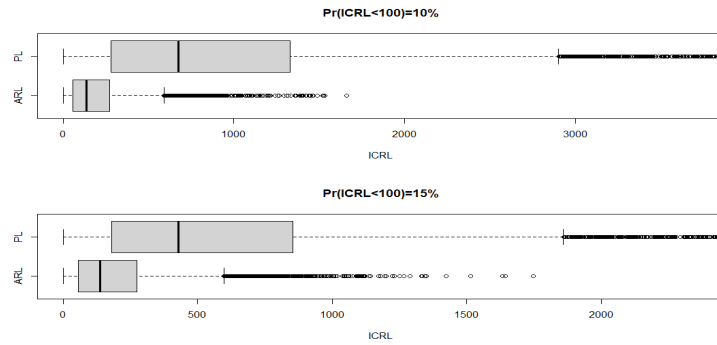
The affine invariance and conditionally distribution-free are the desirable properties of  $T_n$ . Under  $H_0$ ,  $T_n$  has an asymptotic distribution of chi-square random variable with  $p$  degrees of freedom,  $T_n \xrightarrow{d} \chi_p^2$ . Therefore,

$$LCL = 0 \quad \text{and} \quad UCL = \chi_{\alpha}^2(p).$$

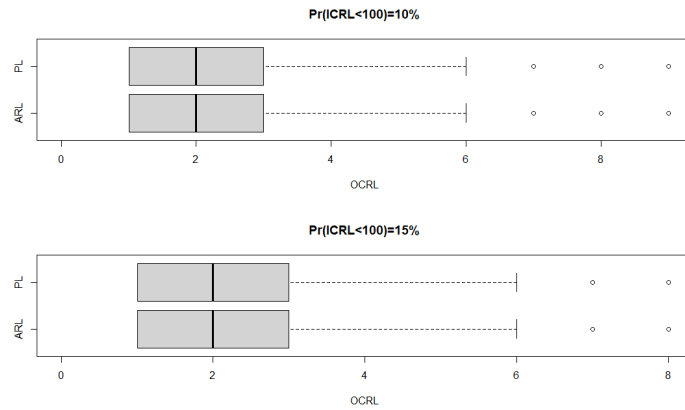
The control chart constants should be adjusted such that  $\Pr(RL_0 > C) = 1 - P_1$ , so we determine the desired type I error rate as  $\alpha = 1 - \exp\left(\frac{\ln(1-P_1)}{C}\right)$  [11].

**3- Findings**

The in-control and out-of-control performances of the ARL  $T_n$  chart are compared with the PL  $T_n$  chart in two different cases by box plots. The results indicate that the PL approach provides just as good performance against the shifts as an ARL chart, and better in-control performance.



**Figure 1-** The box plot of the ICRL ( $RL_0$ ) for the  $T_n$  control chart with PL and ARL approaches for  $n = 50$  and different values of probability level  $P_1$ .



**Figure 2-** The box plot of the OCRL ( $RL_1$ ) for the  $T_n$  control chart with PL and ARL approaches for  $n = 50, d = 3$  and different values of probability level  $P_1$ .

Statistical design of  $T_n$  chart with the percentile-based approach is also provided by solving an optimization problem [11] and [12]. We apply the genetic algorithm (GA) approach to solve the optimization problem such that desired constraints on in-control and out-of-control run lengths are met with prespecified probabilities and we can ensure that the in-control run length exceeds the desired value and the out-of-control run length is less than the desired value.

**4- Discussion and Conclusion**

By comparing the performance of  $T_n$  control chart under the ARL approach with the PL approach, it was found that the performance of the  $T_n$  control chart based on the PL approach is better than the performance of this control chart based on the ARL approach, and in the worst case, is a suitable alternative to the classical ARL approach in statistical design but always an improvement of the classical ARL approach since conceptually, the run lengths are now guaranteed with prespecified probabilities, and are dominant

in terms of run length distributions. Additionally, the PL approach we provide is a generalized form of the MRL and the ARL approaches in the sense that one is not limited to the 50<sup>th</sup> percentile and the average of run length and thus is more flexible than the MRL and ARL approaches.

## References

- [1] Hotelling, H. (1947). *Multivariate quality control*. In: *Techniques of statistical analysis*. McGraw-Hill, New York.
- [2] Montgomery, D. C. (1996). *Introduction to statistical quality control*. John Wiley & Sons Inc., USA.
- [3] Gan, F. F. (2013). *An Optimal Design of EWMA Control Charts Based on Median Run Length*. *Journal of Statistical Computation and Simulation*, 45, 169-184.
- [4] Gan, F. F. (1994). *An Optimal Design of Cumulative Sum Control Chart Based on Median Run Length*. *Communication in Statistics- Simulation and Computation*, 23(2), 485-503.
- [5] Chakraborti, S. (2007). *Run Length Distribution and Percentiles: The Shewhart X Chart with Unknown Parameters*. *Quality Engineering*, 19(2), 119-127.
- [6] Golosnoy, V., & Schmid, W. (2007). *EWMA Control Charts for Monitoring Optimal Portfolio Weights*. *Sequential Analysis (Design Methods and Applications)*, 26(2), 195-224.
- [7] Das, N. (2009). *A Comparison Study of Three Non-parametric Control Charts to Detect Shift in Location Parameters*. *The International Journal of Advanced Manufacturing Technology*, 41(7-8), 799-807.
- [8] Khoo, M. B. C., Wong, V. H., Wu, Z., & Castagliola, P. (2012). *Optimal Design of the Synthetic Chart for the Process Mean Based on Median Run Length*. *IIE Transactions*, 44(9), 765-779.
- [9] Khoo, M. B. C., Wong, V. H., Wu, Z., & Castagliola, P. (2011). *Optimal Design of the Multivariate Synthetic Chart for Monitoring the Process Mean Vector Based on Median Run Length*. *Quality and Reliability Engineering International*, 27(8), 979-1234.
- [10] Dehghan, S., & Faridrohani, M. R. (2018). *Affine Invariant Depth-Based Tests for the Multivariate One-Sample Location Problem*. *TEST*, 28(3), 671-693.
- [11] Faraz, A., Saniga, E., & Montgomery, D. (2019). *Percentile-Based Control Chart Design with an Application to Shewhart  $\bar{X}$  and  $S^2$  Control Charts*. *Quality and Reliability Engineering International*, 35(1), 116-126.
- [12] Haupt, R. L., Haupt, S. E., Haupt S. E. (1998). *Practical Genetic Algorithms*. 2 New York, Wiley.