آزمون آماری ترکیب سمبلها

محمد دخیل علیان
دانشگاه صنعتی اصفهان
دانشکدهٔ برق و کامپیوتر
تلفن: ۸۹۱۲۴۹۳(۰۳۱)
d-alian@cc.iut.ac.ir

محمدرضا عارف
دانشگاه صنعتی شریف
دانشکدهٔ برق
پست الکترونیکی:
mr-aref@cc.iut.ac.ir

بابک صادقیان
دانشگاه صنعتی امیرکبیر
دانشکدهٔ کامپیوتر
تلفن: ۶۴۲۹۴۲۳
basadegh@ce.aku.ac.ar

چکیده:

آزمونهای آماری یکی از ابزارهای ارزیابی دنباله‌های شبه‌تصادفی در سیستمهای رمز پی‌درپی و بسیاری از کاربردهای دیگر می‌باشد. تاکنون آزمونهای متعددی چون فرکانس، سریال، پوکر و... ابداع شده‌اند- که هر یک به نوعی خواص آماری دنباله‌ها را مورد بررسی قرار می‌دهد. در این مقاله ترکیب مولفه‌های زیر قالبهای تشکیل دهندهٔ یک دنبالهٔ شبه‌تصادفی مد نظر قرار گرفته و دو شیوهٔ ترکیب، یکی جمع مولفه‌ها و دیگری پیچیدگی خطی مولفه‌های زیر قالبها مورد بررسی قرارگرفته و در حالت ایده‌آل تابع احتمال هر یک محاسبه شده است. متعاقب آن نیز آزمونهای آماری جدیدی برمبنای این توابع احتمال ارائه شده است.

کلمات کلیدی: آزمونهای آماری، دنباله‌های شبه‌تصادفی، آزمون مربع‌کای، سیستمهای رمز پی‌درپی

۱ مقدمه

ارزیابی آماری عامترین نوع ارزیابی دنباله های شبه تصادفی تولید شده توسط آ لگوریتمهای رمز پی‌درپی می‌باشد -که صرف نظر از نوع ودیدگاه طراحی می‌تواند به کلیه سیستمها اعمال گردد. از آنجا که هدف هر الگوریتم تولید دنباله‌های شبه‌تصادفی است، بنابر این داشتن خواص آماری مطلوب از شرایط لازم برای دنباله‌های تولید شده توسط مولد کلید اجرایی[1] است. در سیستمهای رمز پی‌درپی کلید اجرایی هنگامی مطلوب به حساب می‌آید که هر چه بیشتر به دنباله‌ای با مؤلفه‌های مستقل و با توزیع یکسان (i.i.d) و یکنواخت شبیه‌تر باشد و یا مولد کلید اجرایی نظیر یک منبع باینری بدون حافظهٔ متقارن[2] (B.S.S) عمل کند[۲]. منبع BSS هر یک از دنباله‌های n بیتی ممکن را با احتمالی برابر 2^n تولید می‌کند بنابراین تفاوت قایل شدن میان دنباله‌ها یا به عبارتی تصادفی بودن برخی دنباله‌ها نسبت به برخی دیگر ظاهراً نادرست می‌نماید. برای بررسی میزان تصادفی بودن دنباله‌ها آزمونهای عملی متعددی ابداع شده‌اند. آزمونهای آماری معمولاً بر روی زیر دنباله‌های دنبالهٔ اصلی اعمال می‌شوند، زیرا اولاً بدلیل بزرگ بودن دورهٔ تناوب دنباله‌ها امکان انجام این آزمونها بر روی کل یک دورهٔ تناوب میسر نیست و ثانیاً در یک سیستم پی‌درپی خواص آماری نامطلوب زیر دنباله‌های کلیداجرایی می‌تواند، امنیت سیستم را توسط دشمن مورد تهدید قرار دهد. بنابراین باید هر زیر دنباله از دنبالهٔ کلید اجرایی نیز تصادفی جلوه نماید. آزمونها

1-Running key.

2-Binary symmetric source.

معمولاً بر روی زیر دنباله‌های ۱۰۰۰،۵۰۰، و ... بیشی اعمال شده و در واقع تصادفی بودن آنها به صورت محلی مورد بررسی قرار می‌گیرد. از آنجا که آزمون زیبندگی[1] مربع‌کای[2] معمولاً در اینگونه آزمونها مورد استفاده قرار می‌گیرد، قبل از معرفی آزمون ترکیب سمبلها، توضیح کوتاهی در مورد این آزمون مطرح می‌نماییم.

۲ آزمون زیبندگی مربع کای

نمونه‌های یک متغیر تصادفی با تابع احتمال خاص، نوعاً دارای رفتاری سازگار با تابع احتمال مربوطه می‌باشد. حال اگر نمونه‌های یک متغیر تصادفی در دسترس باشد، چگونه می‌توان بررسی نمود که نمونه‌های مورد نظر با تابع احتمال در نظر گرفته شده برای آن متغیر تصادفی سازگار است؟ آزمون فرضیه[3] بخشی از مبحث آمار و احتمالات است که سعی در پاسخگویی به این سؤال دارد. در آزمون فرضیه یک تابع احتمال به طور فرضی به متغیر تصادفی مورد نظر نسبت داده می‌شود. این آزمون با کمک نمونه‌های متغیر تصادفی، تشخیص می‌دهد که تابع احتمال فرضی برای نمونه‌ها قابل قبول یا مردود می‌باشد. به طور کلی اگر نتایج نمونه‌ها با فرض سازگار به نظر برسد تمایل به پذیرفتن فرض و اگر ناسازگار باشد، تمایل به رد کردن آن فرض پیدا می‌کنیم.

در انجام آزمون دو نوع خطا وجود دارد یکی هنگامی است که فرض را رد می‌کنیم در حالی که واقعاً درست باشد (خطای نوع اول) و دیگری هنگامی است که فرض را قبول می‌کنیم در حالی که نادرست باشد (خطای نوع دوم).بنابراین برای یک فرض اگر بتوان آزمونی یافت که احتمالات ارتکاب هر دو نوع خطا را به طور همزمان به حداقل ممکن برساند، به تعیین بهترین آزمونی خواهد بود که مورد علاقهٔ ماست ولی متأسفانه این کار عموماً دشوار و انجام نشدنی است. خطای نوع اول معمولاً با پارامتر α مشخص می‌گردد و مقدار آن غالباً ۰/۰۵ یا ۰/۱ ..در نظر گرفته می‌شود. در واقع α فاصلهٔ اطمینان مناسبی را جهت انجام آزمون تعریف می‌کند.

در بسیاری از مسایل عملی هدف آزمون کردن دو فرضیه در مقابل یکدیگر می‌باشد. در اینگونه مسایل فرض مشخص بودن تابع احتمال برای یک متغیر تصادفی که نمونه برداری شده است، در مقابل این فرض که تابع احتمال از آن نوع مشخص نباشد مورد آزمون قرار می‌گیرد. یک روش آزمون نمودن چنین فرضهای توزیعی، آزمون زیبندگی مربع‌کای می‌باشد. در همین ارتباط آزمونهای آماری استاندارد نیز معمولاً بر اساس چنین روشی که مبنای آن بکارگیری متغیرهای تصادفی چند جمله ای همراه با نتیجه قضیهٔ زیر است، تعریف شده‌اند.

قضیه۱[۳]:اگر $(X_n,\cdots,X_2,X_1)$ یک متغیر تصادفی چندجمله‌ای با پارامترهای n،P_1،P_2،...،P_m باشد، که در آن X_i هر یک از مقادیر a_j $(j=1,2,\ldots,m)$ را با احتمال زیر بگیرد:

$$p(x_i=a_j)=p_j \qquad j=1,2,\ldots,m \qquad i=1,2,\ldots,n \tag{۱}$$

در این صورت با تعریف x^2_{test} به شکل زیر :

$$x^2_{test}=\sum_{i=1}^{m}\frac{(N_i-nP_i)^2}{nP_i} \tag{۲}$$

هنگامی که $n\to\infty$ میل کند، توزیع متغیر تصادفی x^2_{test} بسمت توزیع مربع‌کای با m-1 درجهٔ آزادی میل خواهد.

N_i در رابطهٔ (۲) تعداد دفعاتی است که a_i در متغیر تصادفی چند جمله‌ای نمونه ظاهر شده است.

برای انجام آزمون مربع‌کای و بررسی این فرض که $(X_n,\cdots,X_2,X_1)$ یک نمونه از متغیر تصادفی چند جمله‌ای با پارامترهای P_1،P_2،...،P_mاست یا خیر، نمونهٔ مورد نظر را اختیار کرده و x^2_{test} را توسط رابطهٔ (۲) محاسبه می‌کنیم. اگر x^2_{test}بزرگتر از $100(1-\alpha)$ امین درصد توزیع مربع‌کای با m-1 درجهٔ آزادی شد($\chi^2_{test}\succ\chi^2_{1-\alpha}$) فرض را رد می‌کنیم. اگر فرض درست باشد احتمال رد شدن نمونه در آن آزمون برابر α می‌باشد. نکته قابل توجه در انجام آزمون بر قرار بودن شرط زیر می‌باشد:

1-Goodness of fit.

2-Chi-square.

3-Hypothesis.

$$\forall i \qquad np_i \succ 5 \qquad i = 1,2,\ldots,m \tag{۳}$$

بسیاری مراجع توافق دارند، تا زمانی که شرط فوق برقرار باشد تقریب توزیع مربع‌کای برای χ^2_{test} مناسب و قابل اعتماد خواهد بود[۴].

۳ آزمون ترکیب سمبلها

همانطور که در بخش قبل ملاحظه نمودیم برای انجام آزمون آماری یک تابع احتمال را به عنوان فرض در نظر می‌گیریم و میزان تطابق دنباله را با مدل مفروض مورد بررسی قرار می‌دهیم. به طرق مختلف و بر حسب نیاز می توان توابع احتمال مورد نظر را تعریف نمود. در آزمونهای متداولی نظیر پوکر[۵] و سریال تعمیم‌یافته[۶]، زیر قالبهای دنبالهٔ مورد آزمون، شمارش شده و پارامتر آزمون محاسبه و با سطح آستانهٔ خاصی مقایسه می‌گردد. در آزمون ترکیب سمبلها به جای شمارش زیر قالبها، ترکیب سمبلهای یک قالب مد نظر قرار می‌گیرد.

فرض کنید دنبالهٔ نمونهٔ مورد نظر به طول KL سمبل، به صورت $S^{KL} = s_1, s_2, \cdots, s_{KL}$ باشد. بنابر این می‌توان آن را به K مولفهٔ L سمبلی تقسیم نمود. حال با ترکیب مولفه‌های هر یک از قالبهای L سمبلی می‌توان تابع احتمال ترکیب مولفه‌ها را به ازای متغیرهای تصادفی مورد نظر جهت انجام آزمون بدست آورد. در ادامه دو ترکیب خطی روی مولفه‌های هر زیر قالب، مورد بررسی قرار می‌گیرد و تابع احتمال هر ترکیب به ازای متغیرهای تصادفی مورد نظر استخراج شده و متعاقب آن پارامتر آزمون مربع‌کای هر یک بیان می‌گردد.

۳-۱ آزمون بر اساس جمع سمبلها

فرض کنید دنبالهٔ S^{KL}، دنباله‌ای باینری با مولفه‌های مستقل و با توزیع یکسان باشد، به طوری که احتمال صفر و یک بودن هر یک از مولفه‌ها برابر p و $q=1-p$ باشد.

$$P(s_i = 1) = p \quad , P(s_i = 0) = q = 1 - p \qquad i = 1,2,\ldots,KL \tag{۴}$$

حال دنبالهٔ $A^K = a_0, a_1, \cdots, a_{K-1}$ که مولفه‌های آن بصورت (۵) تعریف شده است را در نظر می‌گیریم:

$$a_i = s_{iL+1} + s_{iL+2} + \cdots + s_{i(L+1)} \qquad i = 0,1,2,\cdots,K-1 \tag{۵}$$

در این‌صورت دنباله A^K یک دنباله تصادفی با مولفه‌های مستقل از هم می‌باشد. هر یک مولفه‌های A^K یک متغیر تصادفی دو جمله‌ایست که مقادیر $\{0,1,2,\cdots,L\}$ را با احتمالهای زیر اختیار می‌کند:

$$P(a_i = j) = \binom{L}{i} p^j q^{L-j}$$
$$i = 0,1,\cdots,K-1 \qquad , j = 0,1,\cdots,L \tag{۶}$$

با توجه به اینکه دنبالهٔ A^K یک متغیر تصادفی چند جمله‌ای با تابع احتمال(۶) می‌باشد، لذا کلیهٔ شرایط قضیهٔ۱ برای دنبالهٔ A^K برقرار بوده و بکمک آن می‌توان آزمونی را طبق رابطهٔ(۲) تعریف نمود با استفاده از روابط (۲)و(۶) پارامتر جدید χ^2_{cs} را بصورت زیر تعریف می‌نماییم

$$\chi^2_{cs} = \sum_{i=0}^{L} \frac{\left(N_i - KL\binom{L}{i} p^i q^{L-i}\right)^2}{KL\binom{L}{i} p^i q^{L-i}} \tag{۷}$$

در رابطهٔ(۷) N_i تعداد دفعاتی است که i در دنبالهٔ A^K ظاهر شده است

برای انجام آزمون روی دنباله نمونه S^{KL}، ابتدا دنباله را به قالبهای L بیتی تقسیم می‌کنیم وبکمک آن دنبالهٔ A^K را توسط رابطهٔ(۵) بدست آورده وسپس با استفاده از رابطهٔ(۷)، χ^2_{cs} را محاسبه می‌نماییم. اگر مقدار محاسبه شده χ^2_{cs} از سطح آستانه مورد نظر یعنی $\chi^2_{1-\alpha}$ کوچکتر شود، فرض مورد قبول واقع شده وبه عبارت دیگر دنبالهٔ S^{KL} از آزمون عبور خواهد کرد. توجه داشته باشید که سطح آستانه $\chi^2_{1-\alpha}$ بکمک جدول متغیر تصادفی مربع‌کای با L درجهٔ آزادی بدست می‌آید. همچنین برای قابل اعتماد بودن نتیجه آزمون، بر اساس رابطهٔ(۳) باید نامساوی (۸) نیز برقرار باشد.

$$\forall i \in \{0,1,\cdots,L\} \quad ; \quad KL\binom{L}{i}p^i q^{L-i} > 5 \qquad (۸)$$

این آزمون را می‌توان بر روی دنباله‌های کلید اجرایی نیز مورد استفاده قرار داد. در این حالت باید p برابر ۰/۵ اختیار شود. علاوه بر این اگر مولفه‌های دنبالهٔ S^{KL} متعلق به میدان F_q باشند، می‌توان تابع احتمال متغیر تصادفی a_i و متعاقب آن پارامتر آزمون را در این حالت نیز بدست آورد.

ایدهٔ این آزمون می‌تواند، برای بررسی نتایج حاصل شده از آزمونهای دیگر نیز مورد استفاده قرار گیرد. اگر در آزمون آماری T، نتیجهٔ قبول شدن یک دنباله را با یک و رد شدن آن را با صفر نمایش دهیم به عبارت دیگر :

$$T: B^n \rightarrow \{0,1\}$$

(B^n فضای کلیه n بیتی‌های ممکن می‌باشد)

در اینصورت نتایج حاصله از آزمون T برروی n دنبالهٔ مجزا - که توسط مولد شبه‌تصادفی مورد نظر تولید شده‌اند- را بصورت دنباله T^n نمایش می‌دهیم

$$T^n = t_1, t_2, \ldots, t_n \qquad (۹)$$

t_iها یا نتایج آزمون، متغیرهای مستقلی هستند که توزیع هر یک از آنها به مقدار تعریف شده α در آزمون T وابسته است وبصورت زیر بدست می‌آید :

$$P(t_i = 0) = \alpha \quad , P(t_i = 1) = 1-\alpha \qquad i = 1,2,\ldots,n \qquad (۱۰)$$

از آنجا که t_iها مستقل از یکدیگر می‌باشند، طبق قضیه حد مرکزی، T_{cs} که توسط رابطهٔ(۱۱) تعریف می‌شود -با بزرگ شدن n به سمت یک متغیر تصادفی نرمال با میانگین و واریانس $n(1-\alpha)$ و $n\alpha(1-\alpha)$ میل خواهد کرد .

$$T_{cs} = t_1 + t_1 + \cdots + t_n \qquad (۱۱)$$

هنگام انجام آزمونها معمولا سوالی به این شکل مطرح است که از میان n دنبالهٔ مختلف مورد آزمون چه تعداد، نتیجه قبول (یا یک)برای سازگار بودن مولد شبه‌تصادفی با مدل فرضی قبول است؟ اگر فاصله اطمینان را $100(1-\alpha)\%$ در نظر بگیریم تعداد نتایج قبول دنباله‌ها در آزمون -که آن را با N_t نمایش می‌دهیم- با توجه به تقریب توزیع نرمال برای T_{cs} بدست می آید. از آنجا که مجذور یک متغیر تصادفی نرمال معیار[1] یک متغیر تصادفی مربع‌کای می‌باشد. یعنی:

$$\chi^2 = \left(\frac{T_{cs} - n(1-\alpha)}{n\alpha(1-\alpha)}\right)^2 \qquad (۱۲)$$

لذا N_t بایددر محدوده تعیین شده توسط نامساوی (۱۳) قرار گیرد.

$$\left|N_t - n(1-\alpha)\right| < \sqrt{n\alpha(1-\alpha)\chi^2_{1-\alpha}} \qquad (۱۳)$$

به عنوان مثال اگر ۱۰۰۰دنباله شبه‌تصادفی تولید شده توسط یک مولد، مورد آزمون قرار گیرد وفاصله اطمینان ۹۵٪ ($\alpha = 0.05$)در نظر گرفته شود، در صورتی که تعداد دنباله‌های قبول شده در محدودهٔ $936 \le N_t \le 964$ باشد، فرض سازگار بودن توزیع دنباله‌های تولید شده توسط مولد شبه‌تصادفی با تابع احتمال مفروض تایید می‌گردد.

۲-۳ آزمون بر اساس پیچیدگی خطی

فرض کنید دنبالهٔ $S^{KL} = s_1, s_2, \cdots, s_{KL}$ دنباله‌ای با مولفه‌های متعلق به میدان F_q باشد. نظیر آنچه بیان گردید این دنباله را می‌توان به K قالب L مولفه‌ای تقسیم نمود. حال به جای در نظر گرفتن مجموع اعداد یک قالب از رابطهٔ خطی خاصی تحت عنوان پیچیدگی‌خطی که بصورت زیر تعریف می‌گردد، استفاده می‌کنیم.

تعریف : فرض کنید $S^L = s_1, s_2, \ldots, s_L$ دنباله‌ای با مؤلفه‌های باینری و به طول L باشد. در این صورت پیچیدگی‌خطی S^L را

1. Standard.

با $L(S^L)$ نمایش داده و طبق تعریف برابر کوتاهترین LFSR[1]ی است که S^L را تولید کند.

در این‌صورت با استفاده از پیچیدگی‌خطی قالبهای L مولفه‌ای دنبالهٔ S^{KL} می‌توان پارامتر آزمون را بگونه‌ای دیگر نیز تعریف نمود. برای این منظور قضیهٔ زیر را مطرح می‌نماییم.

قضیه ۲: اگر دنبالهٔ S^{KL} دنباله‌ای با مولفه‌های مستقل و با توزیع یکنواخت روی میدان F_q باشد، تابع احتمال پیچیدگی‌خطی قالبهای L مولفه‌ای دنباله دارای تابع احتمال زیر خواهد بود:

$$P\Big(L\big(S^L\big)=i\Big)=\begin{cases} q^{-L} & for\ i=0 \\ (q-1)q^{-min(L+1-2i,2i-L)} & for\ i=1,2,\cdots,L \end{cases} \tag{۱۴}$$

اثبات: از آنجا که دنبالهٔ S^{KL} دنباله‌ای با مولفه‌های یکنواخت است، احتمال $P(L(S^L)=i)$ را می‌توان با توجه فضای نمونهٔ با مدل یکنواخت از طریق شمارش عناصر پیشامد بدست آورد، یعنی:

$$P(L(S^L))=\frac{n(L(S^L)}{n(E)} \tag{۱۵}$$

در رابطهٔ(۱۵)، E فضای نمونهٔ قالبها و $n(.)$ تعداد عناصر پیشامد می‌باشد.

با توجه به اینکه مولفه‌های دنبالهٔ S^{KL}متعلق به میدان F_q است، تعداد اعضای فضای نمونه برابر q^L می‌باشد ($n(E)=q^L$). در [۷] تعداد دنباله‌های L مولفه‌ای که دارای مقدار پیچیدگی‌خطی خاصی هستند به صورت زیر بدست آمده است:

$$n(L(S^L)=i)=\begin{cases} 1 & for\ i=0 \\ (q-1)q^{min(2i-1,2L-2i)} & for\ i=1,2,\cdots,L \end{cases} \tag{۱۶}$$

بنابراین با توجه به رابطهٔ(۱۵) تابع احتمال مورد نظر(رابطهٔ(۱۴)) بسادگی بدست می‌آید.

■

اگر دنبالهٔ $L^K=L_1,L_2,\cdots,L_K$ دنبالهٔ مقادیر پیچیدگی‌خطی قالبهای تشکیل دهندهٔ S^{KL} و در حالتی که دنبالهٔ S^{KL} یک دنبالهٔ با مولفه‌های مستقل و با توزیع یکنواخت باشد، مولفه‌های دنبالهٔ L^K نیز از یکدیگر مستقل بوده و همگی داری توزیع یکسانی با تابع احتمال بیان شده در قضیهٔ۲ خواهند بود. حال با توجه به اینکه کلیهٔ شرایط قضیهٔ۱ برای انجام آزمون برقرار می باشد، پارامتر آزمون جدید را با χ^2_{cL} نمایش داده و به صورت زیر تعریف می‌نماییم.

$$\chi^2_{cL}=\frac{(N_0-KL\times q^{-L})^2}{KL\times q^{-L}}+\sum_{i=1}^{L}\frac{(N_i-KL(q-1)q^{-min(L+1-2i,2i-L)})^2}{KL(q-1)q^{-min(L+1-2i,2i-L)}} \tag{۱۷}$$

(در رابطهٔ(۱۷)، N_iتعداد مولفه‌های با مقدار i در دنبالهٔ L^Kمی‌باشد)

برای انجام آزمون روی دنباله نمونهٔ S^{KL}، ابتدا دنباله را به بلوکهای L بیتی تقسیم می‌کنیم و دنبالهٔ L^K را بدست آورده وسپس با استفاده از رابطهٔ(۱۷)، χ^2_{sL} را محاسبه می‌نماییم. اگر مقدار محاسبه شده χ^2_{sL} از سطح آستانه مورد نظر یعنی $\chi^2_{1-\alpha}$ کوچکتر شود، فرض مورد قبول واقع شده وبه عبارت دیگر دنبالهٔ S^{KL} از آزمون عبور خواهد کرد. توجه داشته باشید که سطح آستانه $\chi^2_{1-\alpha}$نیز بکمک جدول متغیر تصادفی مربع کای با L درجه آزادی بدست می آید. همچنین برای تقریب مناسب توزیع مربع کای، بر اساس رابطه (۳) باید نامساوی زیر نیز برقرارباشد

$$\forall i\in\{0,1,\cdots,L\},\ \ KL\times P\Big(L\big(S^L\big)=i\Big)\succ 5 \tag{۱۸}$$

در سیستمهای رمز پی‌درپی دنباله‌ها به صورت باینری مورد ارزیابی قرار می گیرند (F_2)، لذا در این حالت می‌توان دنبالهٔ L^K را توسط الگوریتمی نظیر برلکمپ-مسی[2] بدست آورد[۸]. در این حالت پارامتر آزمون به صورت رابطهٔ(۱۸) خواهد گردید.

1-Linear Feedback Shift Register.

2-Berlekamp Massey.

$$\chi^2_{cL} = \frac{(N_0 - KL\times 2^{-L})^2}{KL\times 2^{-L}} + \sum_{i=1}^{L} \frac{(N_i - KL\times 2^{-\min(L+1-2i,2i-L)})^2}{KL\times 2^{-\min(L+1-2i,2i-L)}} \qquad (۱۹)$$

از آنجا که توابع احتمال ارائه شده در(۶) و (۱۴) به صورت غیریکنواخت می‌باشند، لذا می‌توان آزمون مورد نظر را به ازای مقادیری که محتمل‌تر هستند انجام داد و بدین‌صورت آزمون را بر روی دنباله‌های کوچکتری اعمال نمود[۹]. در این حالت می‌توان شرایط بیان شده در روابط (۸) و(۱۸) را برای مقادیر مورد نظر در نظر گرفت. به عنوان نمونه اگر $L=۴۰$ باشد، می‌توان مثلاً aهای متعلق به مجموعهٔ $\{11,12,\cdots,29,*\}$ را در نظر گرفت(کلیهٔ مقادیر از ۰ الی ۴۰ که در این مجموعه نیستند به سمبل * نسبت داده شده است). در این حالت A^K یک متغیر تصادفی چند جمله‌ای خواهد بود که مولفه‌های آن مقادیر ۱۱،۱۲،...،۲۹ را با احتمالهای بیان شده در رابطهٔ(۶) اختیار می‌کنند. علاوه بر این داریم:

$$P(a=*) = 1 - \sum_{i=11}^{29} P(a=i) \qquad (۲۰)$$

در این حالت درجهٔ آزادی آزمون به جای ۴۰ برابر ۱۹ خواهد شد. علاوه بر این در این حالت حداقل طول دنباله نیز توسط رابطهٔ(۲۱) محاسبه می‌گردد و بنابراین آزمون را می‌توان به دنبالهٔ کوچکتری نیز اعمال نمود.

$$\forall i \in \{11,12,\cdots,29,*\} \quad KL\times P(a=i) \succ 5 \qquad (۲۱)$$

۴ خلاصه و نتیجه گیری

آزمونهای آماری نقش مهمی را در ارزیابی دنباله‌های شبه‌تصادفی ایفا می‌کنند. در آزمون فرضیه یک تابع احتمال به طور فرضی به متغیر تصادفی مورد نظر نسبت داده می‌شود. این آزمون با کمک نمونه‌های متغیر تصادفی، تشخیص می‌دهد که تابع احتمال فرضی برای نمونه ها قابل قبول یا مردود می‌باشد. در این مقاله ضمن بیان آزمون مربع‌کای در حالت کلی، آزمونی را تحت عنوان "آزمون ترکیب سمبلها" ارائه نمودیم. برای این منظور بر خلاف آزمون پوکر در این ایده ترکیب مولفه‌های زیر قالبهای دنبالهٔ نمونه مد نظر قرار گرفته است. در همین راستا مجموع و پیچیدگی‌خطی مولفه‌ها به عنوان دو شیوهٔ ترکیب مورد بررسی قرار گرفت. بکمک توابع احتمال بدست آمده در این دو روش(به ازای متغیرهای تصادفی در حالت ایده‌آل)، دو آزمون بصورت روابط(۷) و (۱۷) ارائه گردید. غیر یکنواخت بودن این توابع احتمال ما را قادر خواهد نمود تا دنباله‌های با احتمال بیشتر را توسط دنباله‌هایی با طول کوچکتر مورد آزمون قرار دهیم.

مراجع

[۱] دخیل‌علیان، م "ارزیابی دنباله‌های شبه‌تصادفی و طراحی مولدهای آشوبی" ، دانشگاه صنعتی اصفهان، دانشکدهٔ برق و کامپیوتر، رسالهٔ دکترا، آبان ۱۳۷۷.

[2] R.A.Ruepple;*Analysis and Design of Stream Ciphers*,Springer-Verlag, 1986

[3] Larson, G.H.,Introduction to Probability and Statistical Inference, *John-Wiley, 1974*

[4] D.E. Knuth, *The Art of Computer Programming*, Vol.2, Reading MA, Addison Wesley 1981.

[5] H. Beker and F. Piper, *Cipher Systems*, London, Northwood Book,1982.

[6] M. Kimberely, " *Comparison of Two Statistical Test for Keystream Sequences*", Electronic Letter, Vol.23, No.8, pp.365-366, April 1987.

[7] H. Niederreiter, " *The Linear Complexity Profile and the Jump Complexity Keystream Sequences*", Advances in Cryptology, Eurocryp'91, Springer-Verlag, pp.175-189, 1991.

[8] J.L. Maasey, " *Shift Register Synthesis and BCH decoding*", IEEE Trans. on Information Theory vol.15, pp.122-127,1969.