بررسی آزمونهای آماری در سیستمهای رمز پی در پی

دکترمحمد رضا عارف

محمد دخیل علیان

دانشگاه صنعتی اصفهان

چکیده

آزمونهای آماری یکی از ابزارهای مفید برای بررسی تطابق زیر دنباله های کلید اجرایی با دنباله های تصادفی میباشد. در این مقاله ابتدا به بیان تعدادی از مهمترین آزمونهای آماری و چگونگی اعمال این آزمونها بر روی دنباله های نمونه پرداخته میشود. سپس ضمن تجزیه و تحلیل آزمون همبستگی ، دو روش برای اصلاح این آزمون ارائه می گردد. در این روشها تابع خود همبستگی برای دنباله مورد نظرو شیفت یافته های آن محاسبه شده و سپس با توجه به خواص این تابع ، آزمون کای- دوی مناسبی تعریف می شود. در انتهای مقاله آزمون جدید دیگری که مبتنی بر یله ای بودن پیچیدگی خطی دنباله های تصادفی است معرفی می گردد. این آزمون مبتنی بر یک مدل احتمالاتی است که با توجه به خواص پیچیدگی خطی دنباله های کاملاً تصادفی حاصل شده است.

۱ - مقدمه

در سیستمهای رمز پی در پی[1]، دنباله متن اصلی با دنباله اجرایی ، بیت به بیت XOR می گردد تا متن رمز شده را بوجود آورد. از دیدگاه امنیت، حالت ایده آل آن است که دنباله متن رمز شده یک دنباله کاملاً تصادفی باشد. تصادفی بودن در این حالت به معنای استقلال بیتهای دنباله و هم احتمال بودن صفر و یک در دنباله است ($i.i.d$[2]). در این حالت اگر دنباله کلید اجرایی کاملاً تصادفی باشد دنباله رمز شده نیز کاملاً تصادفی باشد. دنباله رمز شده نیز کاملاً تصادفی و غیر قابل پیش بینی خواهد بود شد. برای تولید یک دنباله $i.d.d$ می توان مثلاً با پرتاب سکه بیتهای صفر و یک را تولید کرده و از آن به عنوان دنباله کلید اجرایی استفاده کرد. در این وضعیت باید تمام بیتهای کلید اجرایی مورد استفاده در فرستنده ، عیناً به گیرنده انتقال یابد. بنابراین طول کلید ممکن است بسیار بزرگ شود. بدین لحاظ در عمل از دنباله های شبه تصادفی به عنوان کلید اجرایی استفاده می شود. این دنباله ها از روی یک کلید با طول کوچک تولید می گردند ولی در عین حال دارای دوره تناوب بزرگ و خواص آماری مطلوب می باشند. از دیدگاه کالوب شبه تصادفی بودن در سه قالب تعداد صفرها و یکها و تعداد رنها در یک دوره تناوب مورد نظر را داشته باشند. دنباله های شبه نویزی یا PN[3] نامیده می شوند .

---

[1] *Stream cipher*

[2] *Independently and Identically Distributed*

[3] *Pseudo- random Noise*

هر چه دنباله کلید اجرایی دارای خواص آماری مطلوبتری باشد به عبارت دیگر به دنباله تصادفی نزدیکتر باشد، سیستم از مصونیت بیشتری برخوردار خواهد بود. در عمل آزمونهای متعددی برای بررسی تصادفی بودن دنباله ها مورد استفاده قرار می گیرند. این آزمونها را می توان به دو دسته تقسیم بندی نمود: یکی ، آزمونهای پیچیدگی و دیگری آزمونهای آماری . آزمونهای پیچیدگی با این دیدگاه به دنباله نگاه میکنند که با داشتن چه طولی از دنباله تولید شده می توان کل دنباله را تولید نمود. در حالی که آزمونهای آماری به بررسی میزان تطابق دنباله مورد نظر با یک مدل احتمالاتی می پردازد.

در این مقاله هدف ما بررسی آزمونهای آماری است. بدین لحاظ در بخش دوم به معرفی آزمونهای آماری خواهیم پرداخت و سپس در بخش سوم به طور خاص آزمون خودهمبستگی را مورد بررسی قرار داده و با بیان نقطه ضعف این آزمون دو روش برای اصلاح آن پیشنهاد می گردد. در این دو روش تابع خودهمبستگی برای دنباله مورد نظر و شیفت یافته آن محاسبه و سپس با توجه به خواص آن آزمون زیبندگی [1] کای- دو [2] به نحو مناسبی برروی دنباله اعمال می شود. در بخش چهارم آزمون جدید دیگری که مبتنی بر پیچیدگی خطی دنباله های تصادفی است معرفی می گردد. این آزمون مبتنی بر یک مدل احتمالاتی است که با توجه به خواص نمودار پیچیدگی خطی یک دنباله تصادفی باینری حاصل شده است. آزمون پیچیدگی خطی آزمونی قوی و مناسب برای بررسی خواص آماری دنباله های کلید اجرایی است و نتایج عملی خوبی را نیز به همراه داشته است.

۲ - معرفی آزمونهای آماری

در عمل آزمونهای مختلفی بر روی دنباله ها- برای بررسی تصادفی بودن آنها- صورت می گیرد. آزمونهای آماری استاندارد در واقع رفتار کلی دنباله را با برخی مدلهای خاص احتمالاتی مورد مقایسه قرار داده و میزان تطابق یا عدم تطابق دنباله را با آن مدل تعیین میکنند . آزمونهای آماری معمولاً بر روی زیردنباله های دنباله اصلی اعمال می شوند. زیرا اولاً به دلیل بزرگ بودن دوره تناوب امکان انجام این آزمونها بر روی کل دوره تناوب میسر نیست و ثانیاً در یک سیستم پی در پی نباید به واسطه خواص نامطلوب جزیی از دنباله کلید اجرایی ، سیستم توسط دشمن دچار خدشه شود. بنابراین باید هر زیر دنباله از دنباله کلید اجرایی تصادفی به نظر برسد. آزمونها معمولاً بر روی دنباله های ۵۰۰ ، ۱۰۰۰ و ... بیتی اعمال شده و در واقع تصادفی بودن دنباله ها به صورت محلی مورد بررسی قرار می گیرد. از آنجا که آزمون زیبندگی کای- دو معمولاً در اینگونه آزمونها بکار گرفته می شود قبل از معرفی آزمونها ، توضیح کوتاهی در مورد این آزمون مطرح می نماییم .

۲-۱ آزمون کای- دو

در مبحث متغیرهای تصادفی، توزیعی به نام کای- دو مطرح می باشد. این متغیر تصادفی به این نحو تعریف می گردد که اگر X یک متغیر تصادفی نرمال با میانگین μ و واریانس [2] σ^2 باشد. در این صورت متغیر تصادفی $\frac{(X-\mu)^2}{\sigma^2}$ را یک متغیر تصادفی کای- دو (χ^2) با یک درجه آزادی گویند.

---

[1] *Goodness of fit*

[2] *Chi- Square*

حال اگر n کمیت تصادفی $X_1, X_2, \dots, X_n$ را که مستقل از هم و با توزیع نرمال هستند را در نظر بگیریم . میتوان متغیر تصادفی کای- دو حدیدی به صورت زیر تعریف نمود:

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma_i^2} \qquad (1)$$

تعداد اجزاء تشکیل دهنده تابع $f(X_1, \dots, X_n) = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma_i^2}$ را درجه آزادی متغیر کای- دو میگویند.

قضیه جمع پذیری کای- دو : اگر $X_{n_1}^2, \dots, X_{n_2}^2, X_{n_1}^2$ کمیتهای مستقل از هم باشند به طوری که هر یک به ترتیب دارای درجات آزادی $\nu_1, \dots, \nu_2, \nu_1$ باشند . آنگاه حاصل جمع آنها یعنی :

$$X_\nu^2 = X^2_{\nu_1} + X^2_{\nu_2} + \dots + X^2_{\nu_n}$$

نیز یک متغیر تصادفی کای- دو با درجه آزادی $\nu = \nu_1 + \nu_2 + \dots + \nu_n$ خواهد شد.

در بسیاری از مسایل عملی هدف ما آزمون کردن دو فرض در مقابل یکدیگر می باشد. در اینگونه مسایل فرض مشخص بودن تابع احتمال برای یک متغیر تصادفی که نمونه گیری شده است در مقابل این فرض که تابع احتمال از آن نوع مشخص نباشد . مورد آزمون قرار می گیرد. یک روش آزمون نمودن چنین فرضهایی توزیعی . آزمون زیبندگی کای - دو است . چنین آزمونهایی بر متغیرهای چند جمله ای و قضیه زیر استوار می باشد.

قضیه [۹] : اگر $(X_1, X_2, \dots, X_n)$ یک متغیر تصادفی چند جمله ای با پارامترهای n , $P_1, P_2, \dots, P_m$ باشد، که در آن X_i هر یک از مقادیر a_j $(j = 1,2, \dots, m)$ را با احتمال زیر بگیرد:

$$P(X_i = a_j) = P_j \qquad j = 1,2, \dots, n$$

در این صورت با تعریف U به صورت زیر:

$$U = \sum_{i=1}^{m} \frac{(N_i - nP_i)^2}{nP_i} \qquad (4)$$

هنگامی که $n \to +\infty$ میل کند ، U به سمت توزیع کای - دو با m-1 درجه آزادی میل خواهد کرد. یعنی :

$$\forall k \in R, \quad \lim_{n \to +\infty} F_U(k) = F_{\chi^2}(k)$$

که در آن $F_{\chi^2}(k)$ تابع توزیع متغیر تصادفی کای - دو می باشد که برابر است با :

$$F(k) = \int_0^k \frac{2^{-(\frac{m-1}{2})} y^{(\frac{m-3}{2})} e^{\frac{y}{2}}}{\Gamma(\frac{m-1}{2})} dy$$

N_i در رابطه (٤) تعداد دفعاتی است که a_i در نمونه متغیر تصادفی چند جمله ای ظاهر شده است.

برای انجام آزمون کای- دو و بررسی این فرض که $(X_1, X_2, \dots, X_n)$ یک نمونه از متغیر تصادفی چند جمله ای با پارامترهای معین $P_1, P_2, \dots, P_m$ است یا نه . نمونه مورد نظر را اختیار کرده و U را توسط رابطه (٤) محاسبه می کنیم . اگر $U > \chi^2_{1-\alpha}$، یعنی بزرگتر از $100(1-\alpha)$ امین درصد توزیع کای - دو با m-1 درجه آزادی شد ، فرض را رد می کنیم . اگر فرض درست باشد، احتمال رد شدن آن در آزمون برابر می باشد . در واقع احتمال خطای نوع اول می باشد . تا زمانی که $nP_i > 5$ $(i = 1,2, \dots, m)$ باشد، تقریب کای- دو برای توزیع U کاملاً خوب می باشد. حال با این توضیحات به بیان آزمونهای آماری می پردازیم .

۲-۲ آزمون فرکانس [۱] [۵]

این آزمون اختلاف بین تعداد صفر و یک را در یک دنباله نمونه مورد بررسی قرار می دهد. در یک دنباله باینری تصادفی احتمال وقوع صفر و یک برابر 0.5 می باشد. به عبارت دیگر= P(X=0) = P(X=1) 0.5 می باشد:

$$x_f^2 = \frac{(n_0 - n_1)^2}{n} \qquad (۷)$$

در عبارت فوق n_0 ، n_1 تعداد صفرها و یک ها در دنباله نمونه و n طول دنباله می باشد. در این حالت درجه آزادی برابر یک می باشد. به عنوان مثال اگر 0.5 = در نظر گرفته شود. سطح آستانه مقایسه برای انجام آزمون برابر $x^2_{0.95} = 384$ خواهد شد. اگر x_f^2 برای یک دنباله n بیتی از 3.84 کمتر باشد، دنباله از آزمون قبول خواهد شد و در غیر این صورت رد می گردد.

۲-۳ آزمون سریال [۲] [۵]

این آزمون توزیع دنباله های دو بیتی را در دنباله مورد نظر بررسی می کند. اگر تعداد زیر دنباله های دو بیتی یعنی 00 , 01 , 10 , 11 در یک دنباله نمونه n بیتی به ترتیب برابر n_{00} , n_{01} , n_{10} , n_{11} باشند، می توان نشان داد که در حالت ایده آل (برای دنباله تصادفی نوعی) ، رابطه زیر برقرار خواهد بود:

$$n_{00} = n_{01} = n_{10} = n_{11} \approx \frac{n-1}{4} \qquad (۸)$$

با توجه به این واقعیت پارامتر کای- دو برای این آزمون به صورت زیر تعریف و محاسبه می گردد:

$$x^2_{ser} = \frac{4}{n-1}\sum_{i=1}^{1}\sum_{j=1}^{1}(n_{ij})^2 - \frac{2}{n}\sum_{i=1}^{1}n_i^2 + 1$$

در این حالت درجه آزادی متغیر کای- دو ، برابر ۲ می باشد.

۲-٤ تعداد کل رنها [۳] [۷]

تعداد رنهای موجود در یک دنباله n بیتی دارای توزیعی به شکل نرمال با میانگین زیر است:

$$m = 1 + \frac{2n_0n_1}{n} \qquad (۱۰)$$

$$\sigma^2 = \frac{(m-1)(m-2)}{n-1} \qquad (۱۱)$$

برای انجام آزمون می توان با استفاده از تعریف متغیر تصادفی کای- دو ، پارامتر آزمون را به صورت زیر تعریف نمود:

$$x^2_{run} = (\frac{N_R - m}{\sigma})^2 \qquad (۱۲)$$

در عبارت فوق N_R تعداد کل رنها در دنباله مورد نظر می باشد. کای - دوی فوق دارای یک درجه آزادی است ، زیرا طبق تعریف شامل مجذور تنها یک متغیر تصادفی نرمال استاندارد می باشد.

---

[1] *Frequency test*

[2] *Serial test*

[3] *Run*

۲-۵ آزمون رنها [۵]

در یک دنباله آزمون تصادفی نوعی بر طبق معیار دوم گالوب ، تعداد رنهای به طول I باید در حدود 2^{-i} برابر کل رنها یعنی $n2^{-(i+1)}$ باشد (تعداد کل رنها حدوداً برابر $\frac{n}{2}$ می باشد) لذا با فرض مساوی بودن تعداد گپها[1] و بلوکهای به طول i ، در حالت ایده ال $n2^{-(i+2)}$ گپ به طول i، $n2^{-(i+2)}$ بلوک[2] به طول i خواهیم داشت .

با در نظر گرفتن این مدل پارامتر کای - دو برای گپها و بلوکها به صورت زیر محاسبه می گردد :

$$x_g^2 = \sum_{i=1}^{r} \frac{(r_{gi} - n2^{-(i+2)})^2}{n2^{-(i+2)}} \qquad (۱۳)$$

$$x_b^2 = \sum_{i=1}^{r} \frac{(r_{bi} - n2^{-(i+2)})^2}{n2^{-(i+2)}} \qquad (۱۴)$$

در روابط فوق r_{gi} ، r_{bi} به ترتیب تعداد گپها و تعداد بلوکهای به طول i می باشند. در عمل با توجه به اینکه برای تقریب مناسب کای - دو باید $npi \geq 5$ باشد . لذا باید $r < \log_2(\frac{n}{20})$ باشد. در این حالت درجه آزادی کای - دو برای دو آزمون فوق برابر r خواهد شد.

۲-۶ آزمون پوکر[3] [۵]

برای انجام این آزمون دنباله را به قالبهای m بیتی تقسیم می کنیم . در حالت کلی یک قالب m بیتی می تواند 2^m مقدار مختلف را به خود بگیرد. حال اگر تعداد تکرارهای هر قالب به صورت $f_0, f_1, \ldots, f_{2^m-1}$ بنامیم. در این صورت F را به صورت زیر بیان می کنیم:

$$F = \sum_{i=0}^{2^m-1} f_i = \left\lceil \frac{n}{m} \right\rceil \qquad (۱۵)$$

برای انجام آزمون پوکر پارامتر کای - دو به صورت زیر تعریف می گردد :

$$x_p^2 = \frac{2^m}{F} \sum_{i=0}^{2^m-1} (f_i)^2 - F \qquad (۱۶)$$

این آزمون می تواند به ازای m های مختلف بر روی دنباله اعمال شود. کای - دو فوق دارای $2^m - 1$ درجه آزادی است. لازم به ذکر است که برای انجام آزمون باید $F > 5 \times 2^m$ باشد.

۲-۷ آزمون سریال تعمیم یافته [۸]

آزمون سریال را می توان به زیر دنباله های m بیتی نیز اعمال نمود. برای این کار کلیه دنباله های m بیتی و $m-1$ بیتی را در دنباله شمارش نموده و توسط آن پارامتر کای - دو آزمون به صورت زیر محاسبه می گردد:

$$x_{ser}^2 = \frac{2^m}{n} \sum_{\forall t \in B^m} (nt - n2^{-m})^2 - \frac{2^{m-1}}{n} \sum_{\forall j \in B^{m-1}} (nj - n2^{-m+1})^2 \qquad (۱۷)$$

---

[1] *Gap*

[2] *Block*

[3] *Poker test*

در عبارات فوق B^m و B^{m-1} مجموعه کلیه m و $m-1$ بیتی ها می باشد. کای - دو فوق دارای 2^{m-1} درجه آزادی می باشد. شرط اعمال این آزمون این است که $n > 5 \times 2^m$ باشد. این آزمون نسبت به آزمون پوکر قوی تر بوده و می تواند به دنباله های کوتاهتری نسبت به آن اعمال گردد. البته به جای این آزمون می توان از آزمون پوکر اصلاح شده استفاده نمود یعنی به جای اینکه آزمون پوکر را فقط به دنباله اعمال کنیم، آنرا به شیفت یافته های دنباله نیز اعمال نماییم. این آزمون به آزمون پوکر تعمیم یافته معروف است. بنابراین می توان از آزمون پوکر تعمیم یافته به جای آزمون سریال تعمیم یافته استفاده نمود [۸].

۲-۸ آزمون عمومی[1] ماورر[2] [۱۱]

در این آزمون مولد دنباله تصادفی به صورت یک منبع ارگادیک ایستان[3] با حافظه محدود و با احتمال گذرهای حالت نامعلوم مدلسازی می شود. آزمون، ارتباط نزدیکی با آنتروپی منبع دارد و می تواند اکثر خواص آماری نامطلوب دنباله را تشخیص دهد. اگر مدل منبع ما دارای حافظه M باشد، پارامتر آزمون یعنی L را به صورت $L \geq M$ انتخاب می کنیم. فرض کنید طول دنباله نمونه S^N برابر $N = (Q + K)L$ باشد. K تعداد قدمهای اولیه برای شروع آزمون می باشد. فرض کنید $b_n(S^N)$ بلوک L بیتی n ام از دنباله S^N باشد. یعنی:

$$b_n(S^N) = sL(n-1)+1, sL(n-1)+2, \ldots, sLn \quad \& \quad 1 \leq n \leq Q+K \qquad (۱۸)$$

حال تابع زیر را در نظر می گیریم:

$$f_{T_u}(S^N) = \frac{1}{k}\sum_{n=Q+1}^{Q+K} \log_2(A_n(S^N)) \qquad (۱۹)$$

که در عبارت فوق برای $Q+1 \leq n \leq Q+K$ مقدار $A_n(S^N)$ به صورت زیر بدست می آید:

$$A_n(S^N) = \begin{cases} n & if \forall i \in \{1,2,\ldots,n-1\}: b_n(S^N) \neq b_{n-i}(S^N) \\ \min\{i: i \geq 1, b_n(S^N) = b_{n-i}(S^N)\} & otherwise \end{cases} \qquad (۲۰)$$

برای دنباله باینری R^N با متغیرهای $i.i.d$ میانگین و واریانس $f_{T_u}(R^N)$ به صورت زیر محاسبه شده است:

$$E[f_u(R^N)] \approx L - 0.832746 \qquad L \gg 1 \qquad (۲۱)$$

$$Var[f_u(R^N)] = C(L,K)^2 \frac{Var[\log_2(A_n(R^N))]}{K} \qquad (۲۲)$$

که در عبارت فوق $C(L,K)$ تقریباً برابر است با:

$$C(L,K) \approx 0.7 - \frac{0.8}{L} + (1.6 + \frac{12.8}{L})K^{-\frac{4}{L}}, \quad K \geq 2^L \qquad (۲۳)$$

همچنین داریم:

$$Var[\log_2(A_n(R^N))] \approx 3.423715 \quad for \quad L \gg 1 \qquad (۲۴)$$

در مرجع [۱۱] مقادیر میانگین و واریانس $f_u(R^N)$ به ازای $1 \leq L \leq 16$ در جداولی بیان شده است.

---

[1] *Universal test*

[2] *Maurer*

[3] *Ergodic stationary source*

برای انجام آزمون ماورر می توان پارامتر کای - دوی زیر را تعریف نمود:

$$x^2_{max} = \left(\frac{f_n(S^N) - E\left[f_n(R^N)\right]}{\sigma} \right)^2 \qquad (۲۵)$$

که در رابطه (۲۵) σ برابر:

$$\sigma = C(L,K)\sqrt{\frac{Var\left[\log_2(An(R^N))\right]}{K}} \qquad (۲۶)$$

می باشد. کای- دو مطرح شده در رابطه (۲۵) دارای درجه آزادی یک مـی باشـد. بـرای پیـاده سـازی ایـن آزمون توصیه شده است که پارامتر L در محدوده بیـن 6 الـی 16 و Q بزرگتر از 10×2^L انتخاب شود. همچنین بهتر است که K حتی الامکان خیلی بزرگ انتخاب شود (مثلاً $K = 1000 \times 2^L$).

۳ - بررسی و اصلاح آزمون خود همبستگی [۱]

دنباله های کاملاً تصادفی که مولفه های آن مستقل از هم و با توزیع یکسان باشند . دارای تابع خـود همبستگی غیرهمفاز[1] صفر میباشند. بنابراین در دنباله های نوعی تولید شده توسط یـک منبـع BSS[2] انتظـار داریم تابع خود همبستگی غیر همفـاز مقـدار کوچکی باشـد. تـابع خودهمبسـتگی بـرای یـک دنبالـه متنـاوب $S = \{s_1, s_2, \ldots, s_T, s_1, s_2, \ldots\}$ با دوره تناوب T به صورت زیر تعریف می گردد:

$$C(\tau) = \frac{A-D}{T} = \frac{1}{T}(T - 4\sum_{i=1}^{T} s_i + 4\sum_{i=1}^{T} s_i s_{i+\tau}) \qquad (۲۷)$$

که در عبارت فوق A تعداد بیتهای موافق و D تعداد بیتهای مخالف در دنباله مورد نظر و شیفت یافته آن می باشد. برای دنباله با طول محدود n نظیر $S^n = \{s_1, s_2, \ldots, s_n\}$ تابع خود همبستگی معادل محاسبه تابع خـود همبستگی میان دو دنباله $S_1^{n-\tau} = \{s_1, s_2, \ldots, s_{n-\tau}\}$ و $S_{\tau+1}^{n} = \{s_{\tau+1}, s_{\tau+2}, \ldots, s_n\}$ می باشد که به صورت زیر بدست می آید:

$$C(\tau) = \frac{A-D}{n-\tau} \qquad (۲۸)$$

که در رابطه (۲۸) ، A تعداد بیتهای موافق و D تعدادبیتهای مخالف در دو دنباله است. فرض کنیـد $A(\tau)$ بـه صورت زیر تعریف شود:

$$A(\tau) = \sum_{i=1}^{n-\tau} s_i s_{i+\tau} \qquad (۲۹)$$

در این صورت اگر مولفه های دنباله S^n مستقل از هم باشند . داریم [۵]:

$$\mu_A = E[A(\tau)] = \frac{n_1^2(n-\tau)}{n^2} \qquad (۳۰)$$

در عبارت فوق n_1 تعداد یکها در دنباله S^n می باشد.

[1] *Inphase*

[2] *Binary Symmetric Source*

بر اساس روابط (۲۹) و (۳۰) ، آزمونی مطرح شده است که پارامتر کای - دوی آن به صورت زیر بیان شده است [٤] ، [٥] :

$$x_{ac}^2 = \sum_{\tau=1}^{r} \frac{(A(\tau) - \mu_\tau)^2}{\mu\tau} \qquad (۳۱)$$

اگر دنباله A^1 را به صورت زیر تعریف کنیم :

$$A^\tau = S_1 S_{1+\tau}, S_2 S_{2+\tau}, \ldots, S_{n-\tau} S_n = a_1, a_2, \ldots, a_{n-\tau}, \qquad (۳۲)$$

در این صورت برای یک خاص آزمون کای - دو در صورتی قابل انجام است که مولفه های دنباله A مستقل از هم باشند. در صورت صحت چنین فرضی طبق قضیه حد مرکزی $A(\ )$ برای n های بزرگ با تقریب خوبی دارای توزیع نرمال می باشد . اما چنین فرضی برای دنباله A صحیح نیست. به عنوان نمونه در دنباله A^1 مولفه های مجاور هم مستقل نیستند . زیرا : $(i = 1,2,\ldots,n-1), P(a_i / a_{i-1}) \neq P(a_i)$ میباشد

$$A^1 = S_1 S_2, S_2 S_3, \ldots, S_{n-1} S_n = a_1, a_2, \ldots, a_{n-1} \qquad (۳۳)$$

بنابراین با توجه وابستگی میان مولفه های دنباله A و عدم استقلال این دنباله ها با یکدیگر استفاده از رابطه (۳۱) خالی از اشکال نیست. نتایج عملی این آزمون به وضوح این نقص را نشان می دهد.

برای انجام آزمون خود همبستگی به جای دنباله A^1، دنباله C^τ را به صورت زیر تعریف می کنیم :

$$C^\tau = S_1 \oplus S_{1+\tau}, S_2 \oplus S_{2+\tau}, \ldots, S_{n-\tau} \oplus S_n = C_1^\tau, C_2^\tau, C_{n-\tau}^\tau \qquad (۳٤)$$

(که در عبارت فوق $\oplus$ عملگر XOR می باشد)

با توجه به روابط (۲۸) و (۲۹) تابع خود همبستگی به صورت زیر تبدیل می شود:

$$C(\tau) = \frac{n_0(\tau) - n_1(\tau)}{n-\tau} = 1 - \frac{2n_1(\tau)}{n-\tau} = 1 - \frac{2\sum_{i=1}^{n-\tau} C_i^\tau}{n-\tau} \qquad (۳٥)$$

$\tau = 1,2,\ldots,n-1$

در عبارت (۳۵) ، $n_0(\tau)$ و $n_1(\tau)$ به ترتیب تعداد صفرها و یکها در دنباله C^τ می باشد.

اگر مولفه های دنباله S^n متغیرهای باینری $i.i.d$ باشند در این صورت مولفه های دنباله C نیز $i.i.d$ خواهد شد . [۲]

طبق قضیه حد مرکزی می توان ثابت نمود که با بزرگ شدن n $(n \to +\infty)$ ، متغیر تصادفی $C(\ )$ به سمت یک متغیر تصادفی نرمال با میانگین صفرو واریانس $\frac{1}{n-\tau}$ میل خواهد کرد. بنابراین برای یک خاص می توان پارامتر کای - دو را به صورت زیر تعریف نمود:

$$x_{ac}^2(\tau) = \left(\frac{C(\tau)}{\frac{1}{\sqrt{n-\tau}}} \right)^2 = \frac{(n-\tau-2\sum_{i=1}^{n-\tau} C_i^\tau)^2}{n-\tau} \qquad (۳٦)$$

بنابراین اگر $x_{ac}^2(\tau)$ محاسبه شده برای دنباله نمونه C از $x_{1-\alpha}^2$ بزرگتر باشد ، فرض ناهمبسته بودن دو دنباله $S_{\tau+1}^n$ و $S_1^{n-\tau}$ را رد میکنیم . در غیر اینصورت دنباله S^n از آزمون خودهمبستگی به ازای مورد نظر عبور می کند.

در این آزمون اگر $\alpha = 0.5$ اختیار شود، با استفاده از جدول کای - دو با یک درجه آزادی، مقدار $\chi^2_{0.5}$ برابر ۳/۸٤ خواهد شد. بنابراین دنباله S^n از آزمون خود همبستگی عبور خواهد کرد، هر گاه χ^2_{ac} کوچکتر از ۳/۸٤ باشد.

حال می خواهیم آزمونی ترتیب دهیم که شامل خودهمبستگی یک دنباله به ازای r های مختلف باشد. اگر n به اندازه کافی بزرگ باشد هر یک از متغیرهای $C(1), C(2), \ldots, C(r)$ دارای توزیع نرمال با میانگین صفر و واریانس $\frac{1}{n-1}, \frac{1}{n-2}, \ldots, \frac{1}{n-r}$ می باشند. از طرف دیگر $C(i)$ ها $(i = 1,2,\ldots,r)$ دوبدو مستقل می باشند. به واسطه نرمال بودن هر یک از $C(i)$ ها نتیجه می گیریم که $C(i)$ ها کلاً مستقل نیز میباشند.

بنابراین پارامتر کای - دو زیر را می توان تعریف نمود [۲]

$$\chi^2_{ac} = \frac{C^2(1)}{\sigma^2_{c1}} + \frac{C^2(2)}{\sigma^2_{c2}} + \ldots + \frac{C^2(r)}{\sigma^2_{cr}} = \sum_{i=1}^{r} \frac{C^2(i)}{\sigma^2_{ci}} \qquad (۳۷)$$

(در رابطه (۳۷)، σ^2_{ci} برابر $\frac{1}{n-i}$ می باشد.)

متغیر کای - دو بیان شده برای این آزمون دارای درجه آزادی r می باشد. زیرا $C(i)$ ها $(i=1,2,\ldots,r)$ مستقل از هم می باشند.

از طرف دیگر می توانیم ابتدا متغیر جدیدی به نام Σ_c را به صورت زیر تعریف کنیم :

$$\Sigma_c = C(1)+C(2)+\ldots+C(r) \qquad (۳۸)$$

به عبارت دیگر، در واقع مجموع توابع خودهمبستگی به ازای r های مختلف می باشد. برای n های به مقدار کافی بزرگ $(n \to +\infty)$ به واسطه نرمال بودن $C(i)$ ها و همچنین استقلال آنها از یکدیگر، متغیر تصادفی Σ_c نیز یک متغیر تصادفی نرمال با میانگین صفر و واریانس زیر خواهد شد :

$$\sigma^2_{\Sigma} = \sum_{i=1}^{r} \frac{1}{n-i} \qquad (۳۹)$$

بنابراین پارامتر کای - دوی جدیدی را به صورت زیر می توان تعریف نمود:

$$\chi^2_{ac} = \frac{\Sigma^2_c}{\sigma^2_{c1} + \sigma^2_{c2} + \ldots + \sigma^2_{cr}} = \frac{\left[\sum_{i=1}^{r}\left(1 - \frac{2\sum_{j=1}^{n-i} C_j^i}{n-i}\right)\right]^2}{\sum_{i=1}^{r}\frac{1}{n-i}} \qquad (٤٠)$$

متغیر کای - دو بیان شده در (٤٠) دارای یک درجه آزادی است.

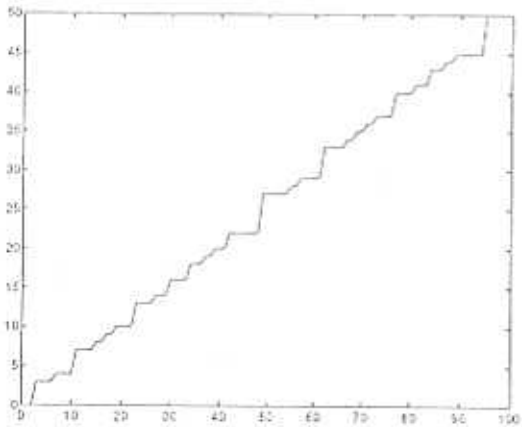
برای انجام آزمون خودهمبستگی بر روی یک دنباله نمونه می توان مقادیر $C(i)$ $(i=1,2,\ldots,r)$ را توسط (۳۵) محاسبه نمود و به کمک آن χ^2_{ac} را محاسبه نمود (به کمک رابطه (۳۷) و (٤٠)). اگر مقدار محاسبه شده χ^2_{ac} از سطح آستانه مورد نظر یعنی $\chi^2_{1-\alpha}$ بزرگتر باشد، فرض ناهمبسته بودن دنباله S^n با شیفتهای خودش رد می شود، در غیر این صورت دنباله S^n از آزمون خود همبستگی عبور خواهد کرد. توجه داشته باشید که سطح آستانه $\chi^2_{1-\alpha}$ بر حسب استفاده از روابط (۳۷) و یا (٤٠) به کمک جدول کای - دو با r درجه آزادی یا یک درجه آزادی باید بدست آید. لذا سطح آستانه این دو رابطه به ازای یک ثابت کاملاً متفاوت خواهد شد.

در عمل با توجه به اینکه در رابطه (۳۷) وزن هر یک از C(i) ها توسط واریانس خودشان مشخص می شود، لذا نقش هر یک از C(i) ها به صورت واقعی تر برجسته میشودو نتایج بهتری را بـه دنبال خواهد داشت . در حالی که در رابطه (٤٠) به نحوی متوسط گیری روی کلیه C(i) ها صورت گرفته است که این امر باعث هموار شدن رابطه χ^2_{obs} شده و نهایتاً تعداد بیشتری از دنباله ها از آزمون عبور خواهند کرد.

٤ آزمون پیچیدگی خطی [۱]

دنباله های شبه تصادفی باید دارای پیچیدگی خطی بزرگ باشند، اما این به تنهایی کافی نیست، چرا که در بسیاری دنباله ها علیرغم بالا بودن پیچیدگی خطی ، فاقد خواص آماری مطلوب مـی باشند. بنابراین علاوه بر بزرگ بودن پیچیدگی خطی باید پیچیدگی خطی به صورت پله ای افزایش یابد .

اگر نمودار پیچیدگی خطی یک دنباله را در مقابل تعداد بیتهای متناظرش ترسیم کنیم منحنی خاصی ایجاد میشود که به آن LCP [۱] میگویند شرط پله ای بودن پیچیدگی خطی به این معناست که LCP دنباله باید به صورت پله ای افزایش یابد. بنابراین دنباله هایی که پیچیدگی خطی آنها به صورت جهشی (و نه پله ای) به حداکثر خود می رسند دارای ضعف می باشند .



شکل (۱) : نمودار LCP یک دنباله تصادفی نمونه

اگر دنباله $S^n = s_1, s_2, \ldots, s_n$ شامل n متغیر باینری مستقل و یکنواخت باشد، میانگین و واریـانس پیچیدگی خطی آن (برای n های بزرگ) به ترتیب برابر $\frac{n}{2}$ و $\frac{86}{81}$ خواهد شد [١٤] . پس برای یک دنبالـه تصادفی نمونه به طول n انتظار این است که پیچیدگی خطی آن بسیار نزدیک به $\frac{n}{2}$ باشـد . از طـرف دیگـر مقدار برش در نمودار LCP دنباله باید به صورت نامنظم به حداکثر خـود برسـد. شکل (۱) نمونـه ای از LCP یک دنباله کاملاً تصادفی که از آزمایش پرتاب سکه حاصل شده است را نشان می دهد. همـانطور کـه ملاحظه میگردد، نمودار کاملاً نامنظم می باشد.

[۱] Linear Complexity Profile

برای استفاده از واقعیت تصادفی بودن پرشها در نمودار LCP می توان قضیه زیر را اثبات نمود:

قضیه [۳] : اگر $S = s_1, s_2, \ldots$ یک دنباله تصادفی باینری با مولفه های $i.i.d$ باشد، آنگاه مدل احتمالاتی برای ارتفاع پرشها در LCP دنباله S^* برابر است با :

$$P(h=m)=\begin{cases}\frac{3}{4} & if \quad m=0\\ \frac{1}{2^{m+2}} & if \quad m=1,2,3,\ldots\end{cases} \qquad (٤١)$$

با توجه به قضیه فوق اگر $H^n = h_1, h_2, \ldots, h_n$ دنباله پرشها در LCP باشد، انتظار داریم به طور متوسط $\frac{3}{4}$ ، $\frac{1}{8}$ ها برابر صفر، $\frac{1}{16}$ آنها برابر یک، آنها برابر دو و همینطور الی آخر. برای انجام آزمون پله ای بودن یا پیچیدگی خطی می توان از این نتایج استفاده نمود. به این ترتیب که یک دنباله شبه تصادفی وقتی از جهت پیچیدگی خطی مناسب است که آمارگان پله های موجود در LCP آن بسیار شبیه رابطه (٤١) باشد.

بر همین اساس و با توجه به رابطه (٤) می توان پارامتر کای - دو به صورت زیر تعریف نمود:

$$x^2_{Lcomp}=\frac{4(Nh0-\frac{3}{4}n)^2}{3n}+\sum_{i=1}^{m}\frac{2^{i-1}(Nhi-\frac{n}{2^{i+2}})^2}{n} \qquad (٤٢)$$

در رابطه (٤٢) N_{hi} تعداد پله های به ارتفاع I در دنباله $H^n = h_1, h_2, \ldots, h_n$ می باشد. شرط انجام آزمون این است که m حداکثر $\log_2(\frac{n}{20})$ انتخاب شود $(m \le \log_2(n/20))$. متغیر کای - دوی بیان شده برای این آزمون دارای درجه آزادی m می باشد.

مثال: فرض کنید دنباله $S^n = s_1, s_2, \ldots, s_n$ توسط رابطه زیر تولید شده باشد:

(٤٣)

$$s=\begin{cases}1 & if \quad n=2^j-1 \quad , j=0,1,2\ldots\\ 0 & otherwise\end{cases}$$

فرض کنید $n = 2^L$ باشد. بنابراین با توجه به اینکه دنباله تولید شده فوق یک دنباله با LCP ایده آل است[۱۳]، داریم :

$$N_h=\begin{cases}2^{L-1} & if \quad i=0,1\\ 0 & otherwise\end{cases} \qquad (٤٤)$$

با توجه به محدودیت بیان شده، m را برابر $L-5$ در نظر می گیریم. بنابراین داریم :

$$x^2_{Lcomp}=\frac{(2^{L-1}-3\times 2^{L-2})^2}{3\times 2^{L-2}}+\frac{(2^{L-1}-2^{L-3})^2}{2^{L-3}}+\sum_{i=2}^{L-5}2^{L-i-2} \succ L-5 \qquad (٤٥)$$

با توجه به رابطه (٤٥) ، اگر $0.05 =$ در نظر گرفته شود ، دنباله فوق هیچگاه از این آزمون عبور نخواهد کرد و این بسیار مطلوب میباشد. نتایج عملی در استفاده از این آزمون نشان داده است که اگر دنباله ای از این آزمون عبور نمود از اکثر آزمونهای مرسوم استاندارد نیز عبور خواهد کرد.

۵ - خلاصه و نتیجه گیری

بکارگیری آزمونهای آماری از جمله قدمهای ارزیابی سیستمهای رمز پی در پی می باشد و به کمک آن می توان به بسیاری از نقایص دنباله کلید اجرایی پی برد. در این راستا در مقاله حاضر به معرفی آزمونهای مختلف آماری پرداخته شد و در مورد هر یک توضیحات مختصری بیان گردید. در یک بیان کلی از میان آزمونهای مطرح شده ، آزمون فرکانس ساده ترین آزمون می باشد - که تنها تفاوت تعداد صفر و یک را در دنباله مد نظر قرار می دهد . آزمون سریال که در واقع تعمیم یافته آزمون فرکانس به دوبیتیهااست ، تنها قادر است وابستگی میان دوبیتیها را در دنباله مدنظر قرار می دهد . واضح است که تعمیم یافته این آزمون و علاوه بر آن آزمون پوکر و تعمیم یافته آن بنحوی در مورد فرکانس بلوکهای بزرگتر و وابستگی بیتها(در حد بلوکها) اظهارنظر می کند . لذا آزمونهای تعمیم یافته از قوت بیشتری برخوردار می باشند ولی در عین حال پیچیده تر بوده و حجم محاسبات و از همه مهمتر به طول بزرگتری از دنباله نیازمند می باشند . آزمون خود همبستگی ارائه شده در این مقاله نیز از این جهت که مستقیماً همبستگی میان بیتهای دنباله را مورد ارزیابی قرار میدهد، قابل توجه و تامل است. خصوصاً هنگامی که از رابطه (۳۷) برای انجام آزمون استفاده شود. قوت این آزمون به مراتب بیشتر میشود و به سادگی می توان با دنباله های نه چندان بزرگ و با حجم عملیات مناسب ، وابستگی بیتهای یک دنباله را مورد ارزیابی قرار داد. مقدار s در این آزمون بهتر است کمتر از $n-10$ انتخاب شود. زیرا در این حالت تقریب کای - دو قابل اعتماد خواهد بود.

از آنجا که آزمون خود همبستگی ارائه شده به نحوی یک آزمون فرکانس است، لذا به نظر میرسد با انتخاب مناسب و بزرگ s می توان از آزمونهای فرکانس ، سریال و پوکر چشم پوشی نمود و تنها به همین آزمون اکتفا نمود.

آزمون عمومی ماورر ، آزمونی قوی است ولی با این وجود نیازمند دنباله های با طول بزرگ و حجم عملیات زیاد است. به عنوان نمونه برای پارامتر $L=16$ طول دنباله باید حدود 10^9 باشد و استفاده از این آزمون تقریباً غیر عملی می گردد. آزمون پیچیدگی خطی ارائه شده در این مقاله ، نیز آزمون بسیار خوبی است که نیاز به دنباله های با طول بزرگ ندارد و به راحتی پیاده سازی می گردد. تنها مساله در این آزمون پیدا نمودن دنباله پرشها در نمودار LCP است که با بکارگیری الگوریتم برلکمپ - مسی [۱۰] قابل انجام است. نکته قابل توجه در این آزمون این است که اگر دنباله ای از این آزمون عبور نمود از اکثر آزمونهای آماری نیز عبور می کند.

۶ - قدردانی

نویسندگان بر خود لازم می دانند از شرکت مهندسی پیام پرداز که انگیزه تهیه مقاله حاضررا فراهم نمود و همچنین به خاطر همکاری این شرکت جهت حروفچینی مقاله تشکر و قدردانی نمایند .

[1] Berlekamp - Massey

۵ - خلاصه و نتیجه گیری

بکارگیری آزمونهای آماری از جمله قدمهای ارزیابی سیستمهای رمز پی در پی می باشد و به کمک آن می توان به بسیاری از نقایص دنباله کلید اجرایی پی برد. در این راستا در مقاله حاضر به معرفی آزمونهای مختلف آماری پرداخته شد و در مورد هر یک توضیحات مختصری بیان گردید. در یک بیان کلی از میان آزمونهای مطرح شده ، آزمون فرکانس ساده ترین آزمون می باشد - که تنها تفاوت تعداد صفر و یک را در دنباله مد نظر قرار می دهد . آزمون سریال که در واقع تعمیم یافته آزمون فرکانس به دوبیتیهااست ، تنها قادر است وابستگی میان دوبیتیها را در دنباله مدنظر قرار می دهد . واضح است که تعمیم یافته این آزمون و علاوه بر آن آزمون پوکر و تعمیم یافته آن بنحوی در مورد فرکانس بلوکهای بزرگتر و وابستگی بیتها(در حد بلوکها) اظهارنظر می کند . لذا آزمونهای تعمیم یافته از قوت بیشتری برخوردار می باشند ولی در عین حال پیچیده تر بوده و حجم محاسبات و از همه مهمتر به طول بزرگتری از دنباله نیازمند می باشند . آزمون خود همبستگی ارائه شده در این مقاله نیز از این جهت که مستقیماً همبستگی میان بیتهای دنباله را مورد ارزیابی قرار میدهد، قابل توجه و تامل است. خصوصاً هنگامی که از رابطه (۳۷) برای انجام آزمون استفاده شود. قوت این آزمون به مراتب بیشتر میشود و به سادگی می توان با دنباله های نه چندان بزرگ و با حجم عملیات مناسب ، وابستگی بیتهای یک دنباله را مورد ارزیابی قرار داد. مقدار τ در این آزمون بهتر است کمتر از $n-10$ انتخاب شود. زیرا در این حالت تقریب کای - دو قابل اعتماد خواهد بود.

از آنجا که آزمون خود همبستگی ارائه شده به نحوی یک آزمون فرکانس است، لذا به نظر میرسد با انتخاب مناسب و بزرگ τ می توان از آزمونهای فرکانس ، سریال و پوکر چشم پوشی نمود و تنها به همین آزمون اکتفا نمود.

آزمون عمومی ماورر ، آزمونی قوی است ولی با این وجود نیازمند دنباله های با طول بزرگ و حجم عملیات زیاد است. به عنوان نمونه برای پارامتر $L=16$ طول دنباله باید حدود 10^9 باشد و استفاده از این آزمون تقریباً غیر عملی می گردد. آزمون پیچیدگی خطی ارائه شده در این مقاله ، نیز آزمون بسیار خوبی است که نیاز به دنباله های با طول بزرگ ندارد و به راحتی پیاده سازی می گردد. تنها مساله در این آزمون پیدا نمودن دنباله پرشها در نمودار LCP است که با بکارگیری الگوریتم برلکمپ - مسی [۱۰][1] قابل انجام است. نکته قابل توجه در این آزمون این است که اگر دنباله ای از این آزمون عبور نمود از اکثر آزمونهای آماری نیز عبور می کند.

۶ - قدردانی

نویسندگان بر خود لازم می دانند از شرکت مهندسی پیام پرداز که انگیزه تهیه مقاله حاضررا فراهم نمود و همچنین به خاطر همکاری این شرکت جهت دروفچینی مقاله تشکر و قدردانی نمایند .

[1] Berlekamp - Massey

۷- مراجع

[۱] محمد دخیل علیان ' طراحی و ارزیابی دنباله های شبه تصادفی غیر خطی در سیستمهای رمز پی در پی ' رساله پتراز دانشگاه صنعتی اصفهان.

برحل ندوین.

[۲] محمد دخیل علیان و محمد رضا عارف ' معرفی یک آزمون خودهمبستگی جدید ' پذیرفته شده در ششمین کنفرانس مهندسی برق ایران *ICEE98* ' دانشگاه خواجه نصیر الدین طوسی ' ۱۳۷۷.

[۳] محمد دخیل علیان ' محمد رضا عارف و محمود مدرس هاشمی ' بررسی پیچیدگی خطی دنباله های تصادفی و یک آزمون آماری ' پذیرفته در ششمین کنفرانس مهندسی برق ایران *CEE98* ' دانشگاه خواجه نصیر الدین طوسی ' ۱۳۷۷.

[۴] محمود مدرس هاشمی ' طراحی سیستمهای رمز کننده پی در پی ' پایان نامه کارشناسی ارشد ' دانشگاه صنعتی اصفهان ' ۱۳۷۰.

[5] H. Beker and F. Piper ; Cipher System: The protection communication, Northwood Books, 1982.

[6] S.W. Golomb ; Shift Register Sequences, Holden-Day, San Fransisco, 1982.

[7] H. Gustfan, Edawson and B. Calli, "Comparison of Block Ciphers", Springer-Werlag, Advances in Cryptology, AUSCRYPT'90 pp.208-219.

[8] M. Kimberely, "Comparison of Two Statistical Test for Keystream Sequences' Electronic Letter, Vol. 23, No. 8, 9th April 1987.

[9] D. Knuth: The Art of Computer Programming, Vol. 2 Addison-Wesley, 1981.

[10] J.L. Massey, "Shift Register Synthesis and BCH Decoding", IEEE Trans. on Information Theory, Vol. 15, No.1 Jan. 1969.

[11] U. Maurer, "A Universal Statistical test for Tandom Bit Generators", Journal of Cryptology, Vol. 5, No. 2, pp. 89-105, 1992.

[12] H. Niederriter, "The Probabilistic Theory of Linear Complexity ", Springer - Verlag, Advances in Cryptology, Eurocrypt'88, pp. 191-210, 1988.

[13] H. Niederriter, "Sequences With almost all Perfect Linear Complexity Profile", Springer - Verlag, Advances in Cryptology, Eurocrypt' 87, pp. 37-52, 1987.

[14] R.A. Rueppel.; Analysis and Design of Stream Cipher; Springer- Verlag, 1986.