آزمون خود همبستگی قالبها

بهروز پی زری(دانشجوی کارشناسی ارشدمخابرات دانشگاه صنعتی اصفهان)
b_peyzari@hotmail.com

محمد دخیل علیان(استادیار دانشگاه صنعتی اصفهان)
mdalian@cc.iut.ac.ir

چکیده:آزمونهای آماری ابزاری جهت بررسی خواص تصادفی دنباله های شبه تصادفی می باشد.در آزمونهای آماری هدف بررسی میزان شباهت دنباله های شبه تصادفی به دنباله ایده ال یعنی دنباله با توزیع یکنواخت که اعضای آن تؤاما مستقل هستند،می باشد.یکی از ویژگی های مهم در بررسی دنباله های شبه تصادفی میزان خود همبستگی اعضای دنباله می باشد.بر همین اساس در[۴] آزمونهای آماری برای سنجش میزان خود همبستگی اعضای دنباله ارائه شده است.یکی دیگر از ویژ گیهای مهم خود همبستگی زیر قالبهای دنباله شبه تصادفی می باشد.در این مقاله با نگرش مطرح شده در[۴]برای خود همبستگی دنباله های باینری ،خود همبستگی قالبها در یک دنباله باینری مورد بررسی دقیق قرار گرفته است و با بدست آوردن مدل احتمال آن از دو دیدگاه آزمونهای آماری جدیدی بر مبنای آن ارائه گردیده است.

واژگان کلیدی:خودهمبستگی ،دنباله های شبه تصادفی،آزمونهای آماری[1]

۱.مقدمه

ارزیابی آماری دنباله های شبه تصادفی در بررسی و تجزیه و تحلیل الگوریتمهای رمزنگاری از اهمیت و جایگاه خاصی بر خوردار می باشند. در واقع دنباله های تولید شده توسط الگوریتمها از دیدگاه تئوری اطلاعات باید از حداکثر بی نظمی و حداکثر آنتروپی بر خوردار باشندو برای نیل به این هدف در یک حالت ایده ال این دنباله ها باید دارای توزیع یکنواخت با اعضای تؤاما مستقل باشندو لی به دلیل اینکه در عمل تولید دنباله های ایده ال میسر نمی باشد مجبور به تولید و استفاده از دنباله های به اصطلاح شبه تصادفی می باشیم. هدف از آزمونهای آماری بررسی میزان نزدیکی دنباله های تولید شده با حالت ایده ال می باشد.در آزمونهای آماری با بدست آوردن مدل احتمال برای ویژگیهایی چون تعداد یکها،تعداد

[1] Statistical Testing

رنها،تعداد بلوکهاو گپها،میزان همبستگی ،میزان پیچیدگی ،میزان فشرده شدن دنباله و... میزان نزدیکی دنباله تولید شده را با حالت ایده ال مورد سنجش قرار می دهندو در نهایت با بدست آوردن آماره آزمون[1] و مقدار احتمال[2] و تعیین سطح معنی داری[3] که توسط آزمون کننده مشخص می شود به تصمیم گیری در مورد دنباله در میزان نزدیکی آن به دنباله ایده ال در خاصیت مذکور مبادرت می شود.در این مقاله با بدست آوردن مدل خود همبستگی قالبها به دو صورت به ارائه آزمونهای آماری جدیدی در این زمینه پرداخته شده است.

۲.مفهوم خود همبستگی

همانطور که در[۴] آمده است میزان خود همبستگی دو دنباله با اعضای باینری با تعداد یکهای موجود در دنباله ای که از XOR عناصر نظیر به نظیر آن دو دنباله حاصل می شود وابسته است.همینطور میزان خود همبستگی دو قالب با اعضای باینری با تعداد یکهای موجود در قالبی که از XOR عناصر نظیر به نظیر آن دو قالب حاصل می شود وابسته است . برای بدست آوردن مدل احتمال این مطلب در ادامه دو قضیه مهم را می آوریم.

۳.دو قضیه اساسی

قضیه۳-۱: قضیه زیبندگی کای دو[۲] : اگر $(X_1, X_2, X_3, \ldots, X_n)$ نمونه های مستقل متغیر تصادفی چند جمله ای X با پارامتر های $n, p_1, p_2, \ldots, p_m$ باشد که در آن X_i $(i = 1,2,\ldots,n)$ هر یک از مقادیر a_j $(j = 1,2,\ldots,m)$ را با احتمال زیر بگیرند:

$$p(X_i = a_j) = p_j \qquad ,j = 1,2,\ldots,m \qquad ,i = 1,2,..,n$$

در این صورت اگر n به اندازه کافی بزرگ باشد عبارت زیر به سمت توزیع کای دو با $m-1$ در جه آزادی میل خواهد کرد.

$$T_n(obs) = \sum_{i=1}^{m} \frac{(N_i - np_i)^2}{np_i} \qquad (۱)$$

که در آن منظور از N_i تعداد دفعاتی است که a_i در متغیرتصادفی چند جمله ای نمونه ظاهر شده است.نکته مهم دیگر در استفاده از این قضیه بر قرار بودن شرط زیر است تا تقریب توزیع کای دو مناسب باشد:

---

[1] Test Statistic
[2] Probability Value
[3] Level of Signficance

$$\forall i \qquad np_i \geq 5 \qquad ,i = 1,2\ldots,m$$

قضیه۳-۲: قضیه خود همبستگی دنباله های $i.i.d$ با توزیع یکنواخت[۴] :اگر $S^n : s_1, s_2, \ldots, s_n$ یک دنباله تصادفی باینری با مؤلفه های تؤاما مستقل و با توزیع یکنواخت باشند،آنگاه دنباله C^τ که به عنوان دنباله خود همبستگی مرتبه τ ام دنباله نا میده می شود نیز دنباله ای با مولفه های تؤاما مستقل و با توزیع یکنواخت می باشند.

$$C^\tau : s_1 \oplus s_{\tau+1}, s_2 \oplus s_{\tau+2}, \ldots, s_{n-\tau} \oplus s_n = c_1^\tau, c_2^\tau, \ldots, c_{n-\tau}^\tau$$
$$\tau = 1,2,\ldots,n-1$$

۴.آزمون خود همبستگی قالبها نوع یک

یکی از ویژگیهای دنباله های خوب این است که بیتها یا قالبهای آنها دارای همبستگی نباشند.با توجه به ایده همبستگی دنباله های باینری که در [۴] آمده و بر اساس آن آزمون خود همبستگی نوع یک و دو ارائه گردیده است، آزمون خود همبستگی قالبها را با بدست آوردن مدل احتمال آنها به صورت زیر ارائه خواهیم کرد.ابتدا از تعاریف مقدماتی زیر شروع می کنیم.

دنباله باینری $S^n : s_1, s_2, \ldots, s_n$ به طول n را به زیر قالبهایی به طول L تقسیم می کنیم.nوLبه گونه ای انتخاب می شوند که $\frac{n}{L} = 2M$ عددی صحیح و زوج باشد.حال $2M$ قالب بدست آمده را به مجمو عه هایی تقسیم می کنیم که اعضای آن از XORشدن متناظر دو قالب بدست آمده باشدو هیچ قالبی بیش از یک بار مورد استفاده قرار نگیرد. مثلا در حالت $M = 2$ فرض کنید که چهار قالب B_1, B_2, B_3, B_4 را داریم. مجموعه های بدست آمده با ویژ گی های بالا به صورت زیر خواهند بود:

$$A_1 = \{B_1 \oplus B_2, B_3 \oplus B_4\}$$
$$A_2 = \{B_1 \oplus B_3, B_2 \oplus B_4\}$$
$$A_3 = \{B_1 \oplus B_4, B_2 \oplus B_3\}$$

که در آن منظور از $B_i \oplus B_j$ عناصر قالبی است که از XOR نظیر به نظیر دو قالب B_i و B_j بدست می آیند. در واقع با این کار همانطور که اشا ره شدهمبستگی دو قالب را مورد سنجش قرار داده ایم. برای ادامه کار دو لم زیر را مطرح می کنیم:

لم۴-۱:اگر s_i, s_j, s_k, s_l متغیرهای تصادفی باینری با توزیع یکنواخت و مستقل باشندآنگاه متغیر های تصا دفی زیر دو به دو مستقل بوده و دارای توزیع یکنواخت می باشند[۴].

$$s_i \oplus s_j, s_l \oplus s_k, s_l \oplus s_j, s_i \oplus s_k, s_j \oplus s_k, s_i \oplus s_l$$

لم۴-۲:با تعریف فوق برای مجموعه ها ،تعداد مجموعه ها در حالت کلی برابر است با:

$$T_c = \frac{(2M)!}{2^M \times (M)!} \qquad (۲)$$

اثبات:چون $2M$ قالب متفاوت داریم این $2M$ قالب در حالت کلی $(2M)!$ جایگشت دارند.در هر جایگشت با انتخاب هر دو قالب و XORآنها به یک قالب جدید می رسیم.از آنجا که ترتیب انتخاب این دو قالب مهم نمی باشد،مقدار $(2M)!$ باید به ازای هر دو قالب بر دو تقسیم شود. چون از XORشدن قالبها Mقالب بدست می آیدپس در نهایت $(2M)!$ بر 2^M تقسیم می شود.از آنجا که جابجایی Mقالب بدست آمده هم در تعریف مجموعه ها مهم نمی باشد،مقدار $\frac{(2M)!}{2^M}$ بر $(M)!$ نیز تقسیم می شود تا تعداد کل مجموعه ها با ویژگیهای بالا بدست آید. ■

ما درادامه عنصرj ام مجموعه i را با a_j^i نشان می دهیم.تعداد اعضای هر مجموعه برابر با ML می باشد.بنابراین مجموعه A_i را در حالت کلی به صورت زیر می توان نوشت:

$$A_i = \{a_1^i, a_2^i, a_3^i, \ldots, a_{ML}^i\}, 1 \le i \le T_c$$

تمام اعضای مجموعه با هم متفاوت می باشندو هیچ عضوی از A_i در مجموعه های دیگر وجود ندارد. بنابراین طبق لم(۴-۱) هر عضو مجموعه از تک تک تمام اعضای مجمو عه های دیگر مستقل است.در ضمن طبق قضیه (۳-۲) تمام اعضای مجموعه A_i تؤاما مستقل می باشند.حال C_i را که برابر مجموع تمام اعضای مجموعه A_i می باشد را به صورت زیر در نظر می گیریم :

$$C_i = \sum_{k=1}^{ML} a_k^i \qquad (۳)$$

حال ثابت می کنیم که C_i ها تو ما مستقل می باشند.چون اعضای مجموعه تؤما مستقل می باشندو دارای توزیع یکنواخت نیز می باشند و C_i ها برابر مجموع آنها می باشد،طبق قضیه حد مرکزی اگر ML به اندازه کافی بزرگ باشد C_i یک متغیر تصادفی گوسی خواهد بودو داریم:

$$E(\mathrm{C_i}) = \mathrm{M} \times \mathrm{L} \times E(a_1^i) = \frac{\mathrm{ML}}{2} \qquad (۴)$$

$$VAR(C_i) = M \times L \times VAR(a_1^i) = \frac{ML}{4} \qquad (۵)$$

حال نشان می دهیم C_i ها مستقل می باشند زیرا نرمال و نا همبسته اند:

$$E(C_i \times C_j) = E\{(\sum_{k=1}^{ML} a_k^i) \times (\sum_{l=1}^{ML} a_l^j)\} = \sum_{k=1}^{ML}\sum_{l=1}^{ML} E(a_k^i \times a_l^j) = \sum_{k=1}^{ML}\sum_{l=1}^{ML} E(a_k^i) \times E(a_l^j) =$$

$$= ML \times ML \times \frac{1}{2} \times \frac{1}{2} = E(C_i) \times E(C_j)$$

بنابراین C_i های بدست آمده به این صورت تؤاما مستقل می باشند.حال برای انجام آزمون به اشکال زیر می توان اقدام نمود:

شکل اول:دیدیم که C_i ها از هم مستقل می باشند.چون مقادیر درمحدوده صفر تا ML می باشند و همچنین C_i ها را می توان مجموع یک سری متغیر تصادفی برنولی $p = \frac{1}{2}$ با پارامتر در نظر گرفت،احتمال اینکه C_i برابر مقدار خاص K باشد برابر است با:

$$p_i = P(C_i = K) = \frac{\binom{ML}{K}}{2^{ML}} \qquad (۶)$$

و می توان آماره آزمون را به صورت زیر تشکیل دادکه با توجه به قضیه(۳-۱) دارای توزیع کای دو خواهد بود.

$$T_n(obs) = \sum_{i=0}^{ML} \frac{(f_i - T_c \times \frac{\binom{ML}{i}}{2^{ML}})^2}{T_c \times \frac{\binom{Ml}{i}}{2^{ML}}} \qquad (۷)$$

که درآن f_i برابر با تعداد C_i هایی است که مقدارشان برابر با i می باشد.آماره آزمون دارای درجه آزادی ML می باشدو در نتیجه مقدار احتمال به صورت زیر بدست می آید:[۱و۵]

$$P\_value = gamma(ML, T_n(obs)) \qquad (۸)$$

که در واقع برابر بامساحت زیرمنحنی تابع چگالی متغیر تصادفی کای دو از $T_n(obs)$ تا بی نهایت است.

شکل دوم:چون C_i ها توما مستقل می باشندو دارای توزیع گوسی می باشندو برای هر کدام داریم:

$$\mu = E(\mathrm{C_i}) = \mathrm{M} \times \mathrm{L} \times E(a_1^i) = \frac{\mathrm{ML}}{2} \qquad (۹)$$

$$\sigma^2 = VAR(C_i) = M \times L \times VAR(a_1^i) = \frac{ML}{4} \qquad (۱۰)$$

و چون مربع یک متغیر تصادفی نرمال ،متغیر تصادفی کای دو با یک درجه آزادی است، می توان آماره آزمون را به صورت زیر نیز تشکیل داد:

$$T_n(obs) = \sum_{i=1}^{T_c} \frac{(C_{obs(i)} - \mu)^2}{\sigma^2} \quad (11)$$

که در آن $C_{obs(i)}$ مقدار مشاهده شده C_i می باشدو چون C_i ها مستقلند، $T_n(obs)$ جمع یک سری متغیر تصادفی کای دو مستقل می باشد بنابراین $T_n(obs)$ یک متغیرتصادفی کای دو با T_c درجه آزادی می باشدو در نهایت مقدار احتمال به صورت زیر بدست خواهد آمد[۱و۵]:

$$P\_value = gamma(T_c, T_n(obs)) \quad (12)$$

شکل سوم:عبارت Cرا به صورت زیر تعریف می کنیم:

$$C = \sum_{i=1}^{T_c} C_i \quad (13)$$

که با توجه به مطالب قبل یک متغیر تصادفی گوسی با پارامتر های زیر می باشد:

$$\mu = E(\mathrm{C}) = T_c \times E(C_1) = T_c \times \frac{\mathrm{ML}}{2} \quad (14)$$

$$\sigma^2 = VAR(C) = T_c \times VAR(C_1) = T_c \times \frac{ML}{4} \quad (15)$$

بنابراین در این حالت آماره آزمون به صورت زیر دارای توزیع کای دو با یک درجه آزادی خواهد بود:

$$T_n(obs) = \frac{(C_{obs} - \mu)^2}{\sigma^2} \quad (16)$$

که در آن C_{obs} مقدارمشاهده شده C برای دنباله مورد آزمون می باشدودر نهایت مقدار احتمال به صورت زیر به راحتی بدست خواهد آمد:

$$P\_value = gamma(1, T_n(obs)) \quad (17)$$

۵.خود همبستگی قالبها نوع دو

همانطورکه درآزمون خود همبستگی قالبها نوع یک ملاحظه گردید شرط ML بزرگ یک شرط اساسی در بدست آوردن توزیع ها بود ولی اگر Lخود به اندازه کافی بزرگ انتخاب شود $(L \geq 10)$ می توان آزمون را به صورت زیر نیز ارائه نمود.دنباله را به M قالب Lبیتی تقسیم می کنیم.از این قالبها به تعداد $\binom{M}{2}$ زوج قالب می توان انتخاب کرد و XOR این قالبها را که میزان خود همبستگی آنها را نشان می دهد را بدست‌آورد.

عناصر قالب i ام را به صورت $(a_1^i, a_2^i, a_3^i, \ldots, a_L^i)$ نشان می دهیم.حال C_i را که برابر با مجموع عناصر قالب i ام است بدست می آوریم:

$$C_i = \sum_{k=1}^{L} a_k^i \qquad (۱۸)$$

چون اعضای مجموع جمع یک سری متغیر تصادفی یکنواخت و توما مستقل می باشند،اگر Lبه اندازه کافی بزرگ باشد، C_i تقریبا متغیر تصادفی نرمال با پارامترهای زیر خواهد بود:

$$E(\mathrm{C_i}) = \mathrm{L} \times E(a_1^i) = \frac{\mathrm{L}}{2} \qquad (۱۹)$$

$$VAR(C_i) = L \times VAR(a_1^i) = \frac{L}{4} \qquad (۲۰)$$

حال مانند قسمت قبل می توان نشان دادکه C_i ها توما مستقل می باشند زیرا دارای توزیع گوسی و نا همبسته می باشند.حال برای انجام آزمون مانند قسمت قبل به سه شکل زیر می توان عمل کرد:

شکل اول:دیدیم که C_i ها از هم مستقل اند.چون مقادیر درمحدوده صفر تا L می باشند و همچنین C_i ها را می توان مجموع یک سری متغیر تصادفی برنولی با پارامتر $p = \frac{1}{2}$ در نظر گرفت،احتمال اینکه C_i برابر مقدار خاص K باشد برابر است با:

$$p_i = P(C_i = K) = \frac{\binom{L}{K}}{2^L} \qquad (۲۱)$$

و می توان آماره آزمون را به صورت زیر تشکیل دادکه با توجه به قضیه(۳-۱) دارای توزیع کای دو خواهد بود.

$$T_n(obs) = \sum_{i=0}^{L} \frac{(f_i - \binom{M}{2} \times \frac{\binom{L}{i}}{2^L})^2}{\binom{M}{2} \times \frac{\binom{L}{i}}{2^L}} \qquad (۲۲)$$

که درآن f_i برابر با تعداد C_i هایی است که مقدارشان برابر با i می باشد.آماره آزمون دارای Lدرجه آزادی است و در نتیجه مقدار احتمال به صورت زیر بدست می آید:

$$P\_value = gamma(L, T_n(obs)) \qquad (۲۳)$$

شکل دوم:چون C_i ها تؤما مستقل می باشندو دارای توزیع گوسی می باشندو برای هر کدام داریم:

$$\mu = E(\mathrm{C_i}) = \mathrm{L} \times E(a_1^i) = \frac{\mathrm{L}}{2} \quad (۲۴)$$

$$\sigma^2 = VAR(C_i) = L \times VAR(a_1^i) = \frac{L}{4} \quad (۲۵)$$

و چون مربع یک متغیر تصادفی نرمال ،متغیر تصادفی کای دو با یک درجه آزادی می باشد، می توان آماره آزمون را به صورت زیر نیز تشکیل داد:

$$T_n(obs) = \sum_{i=1}^{\binom{M}{2}} \frac{(C_{obs(i)} - \mu)^2}{\sigma^2} \quad (۲۶)$$

که در آن $C_{obs(i)}$ مقدار مشاهده شده C_i در مورد دنباله مورد آزمون می باشدو چون C_i ها مستقلند، $T_n(obs)$ جمع یک سری متغیر تصادفی کای دو مستقل می باشد بنابراین $T_n(obs)$ یک متغیر تصادفی کای دو با $\binom{M}{2}$ در جه آزادی می باشدو در نهایت مقدار احتمال به صورت زیر بدست خواهد آمد:

$$P\_value = gamma(\binom{M}{2}, T_n(obs)) \quad (۲۷)$$

شکل سوم:عبارت Cرا به صورت زیر تعریف می کنیم:

$$C = \sum_{i=1}^{\binom{M}{2}} C_i \quad (۲۸)$$

که با توجه به مطالب قبل یک متغیر تصادفی نرمال با پارامتر های زیر می باشد:

$$\mu = E(\mathrm{C}) = \binom{M}{2} \times E(C_1) = \binom{M}{2} \times \frac{\mathrm{L}}{2} \quad (۲۹)$$

$$\sigma^2 = VAR(C) = \binom{M}{2} \times VAR(C_1) = \binom{M}{2} \times \frac{L}{4} \quad (۳۰)$$

بنابراین در این حالت آماره آزمون به صورت زیر دارای توزیع کای دو با یک درجه آزادی خواهد بود:

$$T_n(obs) = \frac{(C_{obs} - \mu)^2}{\sigma^2} \quad (۳۱)$$

که در آن C_{obs} مقدارمشاهده شده C برای دنباله مورد آزمون می باشدودر نهایت مقدار احتمال به صورت زیر به راحتی بدست خواهد آمد:

$$P\_value = gamma(1, T_n(obs)) \quad (۳۲)$$

برای انجام آزمون در هر مورد کافی است با توجه به روابط بالا P_value مربوطه محاسبه گردیده وباسطح معنی داری α $(0.001 \le \alpha \le 0.05)$ مقایسه گردد و در نهایت از دو گزاره زیر برای تصمیم گیری نهایی استفاده شود:[۵،۱]

الف:اگرمقداراحتمال محاسبه شده بزرگتریا مساوی مقدارانتخابی α باشد، از دیدگاه این آزمون دنباله تصادفی تلقی می شود.

ب:اگر مقداراحتمال محاسبه شده کوچکتر از مقدار انتخابی α باشد، از دیدگاه این آزمون دنباله تصادفی تلقی نمی شود.

۶.نتیجه گیری

در این مقاله با نگرشی جدیدی که در[۴] برای خود همبستگی مطرح شده، مدل احتمالی خودهمبستگی قالبها را به دو صورت برای اولین بار بدست آوردیم و با توجه به آن ،آزمونهای آماری خودهمبستگی قالبهای نوع یک و دو را به صورت دقیق بیان نمودیم که از آن می توان برای سنجش خودهمبستگی قالبهای تولیدشده در سیستمهای رمز پی در پی و سیستمهای رمز قالبی استفاده نمود. این سنجش درمورد سیستمهای رمز نگاری قالبی از اهمیت بیشتری برخوردار است.

مراجع

[1]Rukhin A.L and Soto J. and Nechvatal J.,...,"A Statistical Test Suite for Random and Pseudo Random Number Generators for Cryptographic Applications"*NIST Special Publication 800-22*, 15 may ,2001

[2]Rohatgi V.K.,"*Introduction to Probability Theory and Mathematical Statistics*",Johnwiley and sons,1976

[3]Knuth D."*The Art of Computer Programming*",vol. II,Addison Weslley Publishing Company,second edition,1983

[۴]دخیل علیان محمد؛*ارزیابی دنباله های شبه تصادفی و طراحی مولد های آشوبی*؛رساله دکتری در رشته مهندسی برق دانشگاه صنعتی اصفهان،آبان ماه ۱۳۷۷

[5]Rukhin A.L,"Testing Randomness: A Suite of statistical Procedure",*Department of Mathematics and Statistics*,Baltimor, USA,2001

[6]Menzes Alfred," *Handbook of Applied Cryptography*",CRC Press ,1997

[7]Nyberg Kaisa ,"Correlation Theorems in Cryptoanalysis",*Discrete Applied Mathematics*, Elsevier Publication,2001