# شبیه سازی تصادفی

## آزمون مولدهای اعداد تصادفی

۳ مهر ۱۳۹۹

# آزمون مولدهای اعداد تصادفی

- آزمون های نظری/خصوصیات
- آزمون های یکنواختی
- آزمون های استقلال

**تعریف**

دنباله ای از اعداد شبه تصادفی $U_i$ یک دنباله قطعی از اعداد در بازه [۰،۱] است که دارای همان خصوصیات آماری مربوطه به عنوان دنباله ای از اعداد تصادفی است.

- نوع توزیع
- تصادفی بودن (استقلال،سفیدی)

## آزمون های تصادفی/خصوصیات

- آزمون رفتار جهانی (کل دوره ها)
- قضیه های ریاضی
- به طور معمول یکنواختی چندبعدی را بررسی می کند.

# آزمون مولدهای اعداد تصادفی

- آزمون نوع توزیع
  1. آزمون های بصری/طرح ها
  2. آزمون کای دو $(\chi^2)$
  3. آزمون کلموگروف – اسمیرنوف

- آزمون استقلال
  1. آزمون های بصری/طرح ها
  2. آزمون روند
  3. طول مدت آزمون روند
  4. آزمون ضرایب همبستگی

# آزمون معنی داری

- ما مدل (شناخته شده) را فرض می کنیم - فرضیه
- ما یک متغیر تصادفی مشخصه خاص را شناسایی می کنیم - آماره آزمون
- اگر آماره آزمون یک مشاهده غیر طبیعی تحت فرضیه باشد،ما این فرضیه را رد می کنیم.

# عبارات کلیدی

- فرضیه / جایگزین
- آماره آزمون
- سطح معنی داری
- ناحیه پذیرش/بحرانی
- قدرت
- p-value

# توزیع چندجمله ای

- $n$ مورد (آیتم)
- $k$ کلاس
- هر آیتم در کلاس $j$ با احتمال $p_j$ قرار می گیرد
- $X_j$ عدد (تصادفی) آیتم های کلاس $j$ است
- می نویسیم : $X = (X_1, ..., X_2) \sim Mul(n, p_1, ..., p_k)$

بنابراین

$$X_j \sim \mathrm{Bin}(n, p_j)$$

و

$$E(X_j) = np_j$$

و

$$\mathrm{Var}(X_j) = np_j(۱ - p_j)$$

و

$$E\left(\frac{X_j - np_j}{\sqrt{np_j(۱ - p_j)}}\right) = ۰$$

و

$$\mathrm{Var}\left(\frac{X_j - np_j}{\sqrt{np_j(۱ - p_j)}}\right) = ۱$$

در نتیجه

$$\frac{X_j - np_j}{\sqrt{np_j(۱ - p_j)}} \overset{n\to\infty}{\sim} N(۰, ۱)$$

به یاد داریم :

$$\frac{X_j - np_j}{\sqrt{np_j(۱ - p_j)}} \overset{n\to\infty}{\sim} N(۰, ۱)$$

بنابراین

$$\left(\frac{X_j - np_j}{\sqrt{np_j(۱ - p_j)}}\right)^۲ = \frac{(X_j - np_j)^۲}{np_j(۱ - p_j)} \overset{asymp}{\sim} \chi^۲(۱)$$

حال k=۲ را درنظر بگیرید :

$$\frac{(X_۱-np_۱)^۲}{np_۱(۱-p_۱)}=\frac{(X_۱-np_۱)^۲(p_۱+۱-p_۱)}{np_۱(۱-p_۱)}=\frac{(X_۱-np_۱)^۲}{np_۱}+\frac{(X_۱-np_۱)^۲}{n(۱-p_۱)}$$

$$=\frac{(X_۱-np_۱)^۲}{np_۱}+\frac{(X_۱-n-n(p_۱-۱))^۲}{n(۱-p_۱)}=\frac{(X_۱-np_۱)^۲}{np_۱}+\frac{(-X_۲+np_۲)^۲}{np_۲}$$

$$=\frac{(X_۱-np_۱)^۲}{np_۱}+\frac{(X_۲-np_۲)^۲}{np_۲}$$

- آماره $\chi^2$
- اثبات می تواند با استقرا کامل شود.

حالت کلی آماره آزمون به صورت

$$T = \sum_{i=۱}^{n_{classes}} \frac{\left(n_{observed,i} - n_{expected,i}\right)^۲}{n_{expected,i}}$$

می باشد.

- آماره آزمون قرار است با توزیع $\chi^۲$ با درجه آزادی df ارزیابی شود. df به طور کلی $n_{\text{classes}} - ۱ - m$ است که m تعدد پارامترهای تخمین زده شده می باشد.
- توصیه می شود همه گروه ها را به گونه ای انتخاب کنید که $n_{\text{expected,i}} \geq ۵$ باشد.

# آزمون توزیع تست کلموگروف-اسمیرنوف

- تابع توزیع تجربی $F_n(x)$ را با توزیع فرض شده $F(x)$ مقایسه کنید.
- برای پارامترهای شناخته شده، آماره آزمون به $F(x)$ بستگی ندارد.
- قدرت بهتر نسبت به آزمون $\chi^۲$
- هیچ گونه ملاحظاتی برای گروه بندی لازم نیست
- فقط برای توزیع های کاملاً مشخص در نسخه اصلی کار می کند

۲۰ عدد متنوع با توزیع $N(۰,۱)$ :

$-۲.۲۰, -۱.۶۸, -۱.۴۳, -۰.۷۷, -۰.۷۶, -۰.۱۲, ۰.۳۰, ۰.۳۹, ۰.۴۱,$

$۰.۴۴, ۰.۴۴, ۰.۷۱, ۰.۸۵, ۰.۸۷, ۱.۱۵, ۱.۳۷, ۱.۴۱, ۱.۸۱, ۲.۶۵, ۳.۶۹$

$X_i$ ها متغیرهای تصادفی مستقل و هم توزیع می باشند که دارای تابع توزیع $F(x) = P(X \leq x)$ می باشند.
هرکدام به یک تابع تصادفی (ساده) $F_{e,i}(x) = ۱_{\{X_i \leq x\}}$ منجر می شود که به

$$F_e(x) = \frac{۱}{n}\sum_{i=۱}^{n} F_{e,i}(x) = \frac{۱}{n}\sum_{i=۱}^{n} ۱_{\{X_i \leq x\}}$$

$$E(F_e(x)) = E(\frac{۱}{n}\sum_{i=۱}^{n} ۱_{\{X_i \leq x\}}) = \frac{۱}{n}\sum_{i=۱}^{n} E(۱_{\{X_i \leq x\}}) = F(x)$$

$$\mathrm{Var}(F_e(x)) = \frac{۱}{n^۲} nF(x)(۱ - F(x)) = \frac{F(x)G(x)}{n}$$

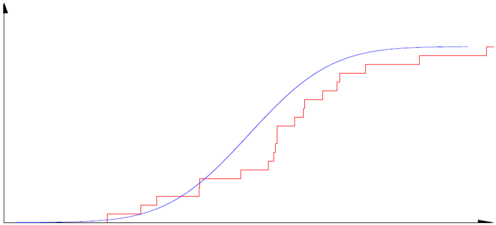
$$F_e(x) \overset{n\to\infty}{\sim} N\left(F(x), \frac{F(x)G(x)}{n}\right)$$

منتهی می شود.

۲۰ عدد متنوع با توزیع N(۰, ۱) :

−۲.۲۰, −۱.۶۸, −۱.۴۳, −۰.۷۷, −۰.۷۶, −۰.۱۲, ۰.۳۰, ۰.۳۹, ۰.۴۱,

۰.۴۴, ۰.۴۴, ۰.۷۱, ۰.۸۵, ۰.۸۷, ۱.۱۵, ۱.۳۷, ۱.۴۱, ۱.۸۱, ۲.۶۵, ۳.۶۹



$$D_n = \sup_x \{|F_n(x) - F(x)|\}$$

آماره آزمون از توزیع کلموگروف پیروی می کند.

| | | Level of significance $(1-\alpha)$ | | | | |
|---|---|---|---|---|---|---|
| Case | Adjusted test statistic | 0.850 | 0.900 | 0.950 | 0.975 | 0.990 |
| All parameters known | $\left(\sqrt{n}+0.12+\frac{0.11}{\sqrt{n}}\right)D_n$ | 1.138 | 1.224 | 1.358 | 1.480 | 1.628 |
| $N(\bar{X}(n), S^2(n))$ | $\left(\sqrt{n}-0.01+\frac{0.85}{\sqrt{n}}\right)D_n$ | 0.775 | 0.819 | 0.895 | 0.955 | 1.035 |
| $\exp(\bar{X}(n))$ | $\left(\sqrt{n}+0.26+\frac{0.5}{\sqrt{n}}\right)\left(D_n-\frac{0.2}{n}\right)$ | 0.926 | 0.990 | 1.094 | 1.190 | 1.308 |

- در آزمون نسخه دو بعدی برای یکنواختی $(U_{2i-1}, U_{2i})$
- به طور معمول در آزمون $\chi^2$
- تعداد گروه ها به شدت با بعد افزایش می یابد

بزرگتر/کوچکتر :

- آزمون روند داده شده می تواند در مقایسه با میانه مورد استفاده قرار بگیرد.
- تعداد روندها (بزرگتر/کوچکتر از میانه) به صورت زیر توزیع شده است :

$$N\left(2\frac{n_1 n_2}{n_1+n_2}+1,\ 2\frac{n_1 n_2(2n_1 n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}\right)$$

که $n_1$ تعداد نمونه های بزرگتر و $n_2$ تعداد نمونه های کوچکتر است.

- آماره آزمون برابر با کل تعداد روندها یعنی $T=R_a+R_b$ است که $R_a$ بالای روندها و $R_b$ زیر روندها است.

افزایشی/کاهشی (knuth) : یک آزمون به طور خاص برای آزمایش مولد اعداد تصادفی طراحی شده است که از آزمون روند افزایشی/کاهشی تبعیت می کند.
دنباله زیر:

$$۰/۵۴, ۰/۶۷, |۰/۱۳, ۰/۸۹, |۰/۳۳, ۰/۴۵, ۰/۹۰, |$$
$$۰/۰۱, ۰/۴۵, ۰/۷۶, ۰/۸۲, |۰/۲۴, |۰/۱۷$$

روندهایی به طول ۲, ۲, ۳, ۴, ۱,... دارد. یعنی روندهای اعدادی که به طور مکرر در حال افزایش هستند.

n عدد تصادفی تولید کنید. عدد مشاهده شده روندهای به طول ۶ ≤ ,۵ ,... ,۱ در بردار R ثبت شده است.آماره آزمون به صورت زیر محاسبه می شود :

$$Z = \frac{1}{n-6}(\boldsymbol{R} - n\boldsymbol{B})^T A(\boldsymbol{R} - n\boldsymbol{B})$$

$$A = \begin{bmatrix} 4529.4 & 9044.9 & 13568 & 18091 & 22615 & 27892 \\ 9044.9 & 18097 & 27139 & 36187 & 45234 & 55789 \\ 13568 & 27139 & 40721 & 54281 & 67852 & 83685 \\ 18091 & 36187 & 54281 & 72414 & 90470 & 111580 \\ 22615 & 45234 & 67852 & 90470 & 113262 & 139476 \\ 27892 & 55789 & 83685 & 111580 & 139476 & 172860 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{24} \\ \frac{11}{120} \\ \frac{19}{720} \\ \frac{29}{5040} \\ \frac{1}{840} \end{bmatrix}$$

و آماره آزمون با توزیع $\chi^{۲}(۶)$ می شود. باید n > ۴۰۰۰ باشد.

آزمون افزایشی کاهشی :
دنباله

۰/۵۴, ۰/۶۷, |۰/۱۳, ۰/۸۹, |۰/۳۳, ۰/۴۵, ۰/۹۰, |
۰/۰۱, ۰/۴۵, ۰/۷۶, ۰/۸۲, |۰/۲۴, |۰/۱۷

به صورت زیر تبدیل می شود:

$$<, >, <, >, <, <, >, <, <, <, >, >$$

در کل ۸ روند به طول های ۲ ,۳ ,۱ ,۲ ,۱ ,۱ ,۱ ,۱ می دهد.

عدد مورد انتظار رروندهای به طول $k$ برای روندهای به طول ۱ و ۲ به ترتیب برابر با $\frac{n+۱}{۱۲}$ و $\frac{۱۱n-۴}{۱۲}$ و برای روندهای به طول $k < N - ۱$ برابر با

$$\frac{۲[(k^۲ + ۳k + ۱)n - (k^۳ + ۳k^۲ - k - ۴)]}{(k + ۳)!}$$

می باشد.

X را برابر با عدد کل روندها تعریف می‌کنیم. آنگاه :

$$Z = \frac{X - \frac{۲n-۱}{۳}}{\sqrt{\frac{۱۶n-۲۹}{۹۰}}}$$

دارای توزیع $N(۰, ۱)$ می باشد.

- همبستگی تخمین زده شده به صورت زیر می باشد :

$$c_h = \frac{۱}{n-h}\sum_{i}^{n-h} U_i U_{i+h} \sim N\left(۰/۲۵, \frac{۷}{۱۴۴n}\right)$$