НЕПРЕКИДНЕ РАСПОДЕЛЕ

У овом делу су дате дефиниције и основе чињенице које се односе на расподеле непрекидних случајних променљивих или краће, *непрекидне расподеле*.

1 Случајне променљиве и функције расподела

Појам случајне променљиве увео је Карл Пирсон[1] 1909. године.

Нека је $(\Omega, \mathcal{F}, P)$ простор вероватноће (Ω је скуп елементарних догађаја, $\mathcal{F}$ је σ-алгебра догађаја, а P је вероватноћа дефинисана на $(\Omega, \mathcal{F})$) и нека је $\mathcal{B}$ Борелова σ-алгебра подскупова скупа реалних бројева R.

Дефиниција 1 *Функција $X : \Omega \to R$ је случајна променљива (величина) ако је мерљива у односу на σ-алгебре $\mathcal{F}$ и $\mathcal{B}$, односно ако за сваки Борелов скуп $B \in \mathcal{B}$ важи*

$$X^{-1}(B) = \{\omega : X(\Omega) \in B\} \in \mathcal{F}.$$

Из дефиниције следи да се на природан начин може дефинисати вероватноћа скупа B за сваки Борелов акуп $B \in \mathcal{B}$. Функција $P_X : \mathcal{B} \to R$ дефинисана са

$$P_X(B) = P(X^{-1}(B))$$

је вероватноћа и назива се расподела вероватноћа случајне променљиве X. Специјално, за $x \in R$ можемо да узмемо $B = (-\infty, x]$.

1.1 Функција расподеле

Дефиниција 2 *Функција $F_X : R \to R$ дефинисана са*

$$F_X(x) = P\{\omega : X(\omega) \le x\}$$

је функција расподеле вероватноћа или краће, функција расподеле случајне променљиве X.

Ако је јасно о којој променљивој је реч, уместо F_X може се писати F. Функција F је растућа, непрекидна с десне стране за сваки реалан број x и важи:

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1.$$

Издвајају се три типа функција расподеле: дискретне, апсолутно непрекидне и сингуларне, при чему свака фунција расподеле F може да се прикаже у облику

$$F(x) = \alpha F_1(x) + \beta F_2(x) + (1 - \alpha - \beta) F_3(x), \quad \alpha, \beta \ge 0,$$

[1] Karl Pearson (1857-1936) - енглески математичар

где је F_1 апсолутно непрекидна, F_2 дискретна и F_3 сингуларна фунција расподеле.

Овде дајемо само дефиницију апсолутно непрекидне функције расподеле обзиром да су у *Атласу* заступљене само непрекидне расподеле.

Дефиниција 3 *Функција расподеле F случајне променљиве X је апсолутно непрекидна ако постоји интеграбилна функција $g : R \to R$ таква да за сваки реалан број x важи*

$$F(x) = \int_{-\infty}^{x} g(t)dt.$$

Функција g је густина расподеле случајне променљиве X.

Апсолутно непрекидна функција расподеле је непрекидна и диференцијабилна скоро свуда, при чему у тачки диференцијабилности x важи $g(x) = F'(x)$. Свака ненегативна интеграбилна функција $g : R \to R$ за коју је $\int_{-\infty}^{+\infty} g(t)dt = 1$ је густина неке расподеле.

1.2 Независност случајних величина

За две или више случајних променљивих дефинише се појам независности.

Дефиниција 4 *Случајне променљиве X и Y на истом простору вероватноће $(\Omega, \mathcal{F}, P)$ су независне ако за свака два Борелова скупа A и B из $\mathcal{B}$ важи*

$$P\{\omega : X(\omega) \in A, Y(\omega) \in B\} = P\{\omega : X(\omega) \in A\} \cdot P\{\omega : Y(\omega) \in B\}.$$

На сличан начин се дефинише независност три или више случајних променљивих. Независност може да се изрази и помоћу функција расподеле. На пример, случајне променљиве X и Y су независне ако и само ако је

$$P\{\omega : X(\omega) \leq x, Y(\omega) \leq y\} = F_X(x) \cdot F_Y(y)$$

за свака два реална броја x и y.

2 Непрекидне случајне променљиве

За разлику од дискретних случајних величина, непрекидна случајна променљива узима вредности из неког интервала или уније интервала реалне праве. При томе, интервал може да буде коначан или бесконачан, као и затворен, отворен или полузатворен.

Дефиниција 5 *Случајна променљива X је апсолутно непрекидна ако је њена функција расподеле F апсолутно непрекидна.*

Ако је случајна променљива апсолутно непрекидна, кажемо да она има апсолутно непрекидну или краће, непрекидну расподелу. Обзиром да фукција расподеле у том случају може да се изрази помоћу густине, важи следеће тврђење.

Теорема 1 *Вероватноћа да вредност апсолутно непрекидне случајне променљиве X припада било ком од скупова (a, b), $(a, b]$, $[a, b)$ или $[a, b]$ једнака је $\int_a^b g(x)dx$.*

У овом *Атласу* заступљене су непрекидне расподеле које имају непрекидну густину. У том случају је функција расподеле диференцијабилна, при чему је $g(x) = F'(x)$ за сваки реалан број x. Из својства 1. наведеног тврђења следи да је

$$P\{\omega : x \leq X(\omega) \leq x + \Delta x\} = g(x)\Delta x + o(\Delta x)$$

када $\Delta x \to 0$.

3 Тип расподеле

Дефиниција 6 *Две случајне променљиве су једнаке ако имају исту функцију расподеле.*

У смислу ове дефиниције, ознака $X = Y$ значи да је

$$F_X(x) = P\{\omega : X(\omega) \leq x\} = P\{\omega : Y(\omega) \leq x\} = F_Y(x)$$

за сваки реалан број x.

Дефиниција 7 *За случајне променљиве X и Y кажемо да имају исти тип расподеле ако постоје реални бројеви a и b $(b > 0)$ такви да је*

$$Y = a + bX.$$

Према дефиницији имамо да је

$$F_Y(x) = P\{Y \leq x\} = P\left\{X \leq \frac{x-a}{b}\right\} = P_X\left(\frac{x-a}{b}\right)$$

за сваки реалан број x. То значи да се функција расподеле за Y може добити помоћу функције расподеле за X линеарном трансформацијом аргумента. Другим речима, избором функције расподеле F одређена је и цела једна двопараметарска фамилија функција расподеле дефинисана са

$$F(x; a, b) = F\left(\frac{x-a}{b}\right), \quad x \in R,\ a \in R,\ b > 0.$$

При томе је $F(x) = F(x; 0, 1)$. Све расподеле ове фамилије су истог типа. Параметар a је *параметар локације*, а параметар b је *параметар скалирања*.

За непрекидне расподеле наведеној фамилији функција расподеле одговара двопараметарска фамилија густина дефинисана са

$$g(x; a, b) = \frac{1}{b} g\left(\frac{x-a}{b}\right),$$

где је g густина расподеле за X. При томе је

$$g_Y(x) = g(x; a, b), \quad g_X(x) = g(x; 0, 1).$$

4 Фамилије расподела

Постоје и фамилије расподела које садрже више типова расподела, при чему све расподеле из фамилије имају неко заједничко својство. Овде се наводе две такве фамилије расподела: *Пирсонове расподеле* и *експоненцијалне расподеле*. У случају фамилије експоненцијалних расподела густине расподела имају исти облик, а у случају Пирсонових расподела густине расподела су решења диференцијалних једначина истог облика.

4.1 Пирсонове расподеле

Развијајући математичку теорију еволуције Пирсон је открио једноставну диференцијалну једначину

$$\frac{g'}{g} = \frac{x-a}{b_0 + b_1 x + b_2 x^2}$$

чија су решења густине дванаест типова расподела. Решења ове једначине се класификују зависно од природе нула полинома $b_0 + b_1 x + b_2 x^2$. На тај начин се добија дванаест типова решења која одређују одговарајуће типове расподела познатих као Пирсонове расподеле (тип I до тип XII).

4.2 Експоненцијалне расподеле

Густине расподела ове фамилије су експоненцијалне функције дате са

$$g(x) = \exp\{a(\theta)b(x) + c(\theta) + d(x)\},$$

где су a, b, c, d реалне функције, а θ параметар.

Специјални случај је фамилија ексоненцијално степених расподела коју је први описао Суботин (Subbotin, 1923). По угледу на Пирсонове расподеле, Лунета (Lunetta, 1963) је експоненцијално степене расподеле разматрао као расподеле чије су густине решења диференцијалне једначине

$$\frac{g'}{g} = k \cdot \frac{\ln g - \ln a}{x - c}.$$

На тај начин се добија тропараметарска фамилија густина

$$g(x; p, \mu, \sigma) = \frac{1}{2\sigma p^{1/p}\Gamma(1 + 1/p)} \exp\left\{-\frac{|x-\mu|^p}{p\sigma^p}\right\},$$

где параметри p, μ и σ зависе од a, c и k.

5 Основне нумеричке карактеристике

Неке нумеричке карактеристике, као што су очекивање, варијанса, модус и медијана, дају основне информације о расподели случајне променљиве.

5.1 Математичко очекивање

Основна мера централне тенденције случајне променљиве је њена очекивана вредност.

Дефиниција 8 *Математичко очекивање $E(X)$ непрекидне случајне променљиве X чија је густина расподеле g је дато са*

$$E(X) = \int_{-\infty}^{+\infty} xg(x)dx$$

под условом да наведени интеграл апсолутно конвергира. Ако то није случај, математичко очекивање не постоји.

За математичко очекивање линеарне комбинације и производа две случајне променљиве важи следеће тврђење.

Теорема 2 *Ако су X и Y случајне променљиве које имају математичко очекивање, тада је:*

1. $E(aX + bY) = aE(X) + bE(Y)$ *за* $a, b \in R$,
2. $E(XY) = E(X)E(Y)$ *ако су* X *и* Y *независне.*

5.2 Модус

Још једна мера централне тенденције је *модус* или *мод*.

Дефиниција 9 *За непрекидну случајну променљиву мод је тачка у којој функција густине достиже локални максимум.*

Према томе, потенцијалне вредности мода су нуле првог извода функције густине. Мод не мора да постоји, а ако расподела има само један мод каже се да је унимодална. Расподела са два мода је бимодална, а са више модова је вишемодална.

Термин *мод* је увео Пирсон 1895. године.

5.3 Дисперзија

Дисперзија или варијанса представља очекивано средњеквадратно одступање од очекиване вредности случајне променљиве.

Дефиниција 10 *Дисперзија $D(X)$ или варијанса $V(X)$ случајне променљиве X је дата са*

$$D(X) = E(X - E(X))^2$$

под условом да наведена очекивања постоје. У противном дисперзија не постоји. Позитивна вредност квадратног корена дисперзије назива се стандардна девијација и означава са $\sigma(X)$ или само σ.

Из дефиниције следи да је

$$D(X) = E(X^2) - (E(X))^2,$$

што значи да је $D(X) \le E(X^2)$ и $|E(X)| \le \sqrt{E(X^2)}$. За дисперзију линеарне комбинације две променљиве важи следеће тврђење.

Теорема 3 *Ако су X и Y назависне случајне променљиве које имају дисперзије и ако су a и b реални бројеви, тада је*

$$D(aX + bY) = a^2 D(X) + b^2 D(Y).$$

Ако је $X^* = \dfrac{X - E(X)}{\sqrt{D(X)}}$ (стандардизовани или нормирани облик случајне променљиве X), тада је $D(X^*) = 1$.

Појам стандардне девијације увео је Пирсон 1894. године, а појам варијансе увео је Фишер[2] 1918. године.

5.4 Коефицијент варијације

Још једна мера одступања од очекивање вредности је *коефицијент варијације* који је такође увео Пирсон.

Дефиниција 11 *Ако је m очекивање, а σ стандардна девијација случајне променљиве X, коефицијент варијације C_V је дат са*

$$C_V = 100\frac{\sigma}{m}.$$

Према томе, овај коефицијент представља количник стандардне девијације и математичког очекивања изражен у процентима.

5.5 Квантили

Обзиром да математичко очекивање и варијанса не морају постојати за дату расподелу, дефинишу се и неке мере централне тенденције и дисперзије које постоје за сваку расподелу.

Дефиниција 12 *За непрекидну случајну променљиву квантил реда или нивоа α $(0 < \alpha < 1)$ је број x_α за који је*

$$F(x_\alpha) = \alpha.$$

Квантили реда $\alpha = 0.25$ и $\alpha = 0.75$ називају се први и трећи квартил и означавају са Q_1 и Q_3. Као мера дисперзије расподеле често се, у случају једнозначно одређених квартила, користи њихова разлика (интерквартилна разлика)

$$Q = Q_3 - Q_1.$$

[2]Ronald Aymler Fisher (1890-1962) – енглески статистичар

5.6 Медијана

Дефиниција 13 *За непрекидну случајну променљиву X квантил реда $1/2$ назива се медијана и означава са $Me(X)$.*

Ако је функција густине симетрична у односу на неку тачку, та тачка је управо медијана, при чему су математичко очекивање и медијана једнаки (уколико очекивање постоји).

6 Моменти

Додатне информације о расподелама се могу добити из вредности момената и централних момената.

Дефиниција 14 *Ако је g функција густине случајне променљиве X, моменат m_r реда r је дефинисан са*

$$m_r = E(X^r) = \int_{-\infty}^{+\infty} x^r g(x) dx,$$

а централни моменат μ_r реда r је дефинисан са

$$\mu_r = E(X-\mu)^r = \int_{-\infty}^{+\infty} (x-m)^r g(x) dx,$$

при чему се претпоставља да су наведени интеграли апсолутно конвергентни. Ако су наведени интеграли дивергентни, одговарајући моменти не постоје.

Према дефиницији, математичко очекивање је моменат првог реда, а дисперзија је централни моменат другог реда,

$$E(X) = m_1, \quad D(X) = \mu_2.$$

Коришћењем биномног развоја за $(x-m)^r$ могу се добити везе између момената и централних момената. Наиме, важе једнакости

$$\mu_k = \sum_{i=0}^{k} (-1)^i \binom{k}{i} m_{k-i} m^i,$$

за $k \in N$, где је $\mu_0 = m_0 = 1$. Слично се добијају и једнакости

$$m_k = \sum_{i=0}^{k} \binom{k}{i} \mu_{k-1} m^i, \quad k \in N.$$

7 Мере асиметрије и спљоштености

Поред основних нумеричких карактеристика дате расподеле, некада су важни и подаци који се односе на облик расподеле. Најзначајнији међу њима су они који говоре о симетрији, односно асиметрији и о спљоштености расподеле, а познати су као *Пирсонови* коефицијенти.

Дефиниција 15 *Расподела непрекидне случајне промењиве X је симетрична уколико постоји тачка a за коју је*

$$g(a-x) = g(a+x)$$

за сваки реалан број x. У супротном расподела је асиметрична.

За асиметричне расподеле потребна је и мера асиметрије.

Дефиниција 16 *Коефицијент асиметрија $\pi_1(X)$ је дефинисан са*

$$\pi_1(X) = \frac{\mu_3}{\mu_2^{3/2}}$$

За $\pi_1(X) = 0$ расподела је симетрична, за $\pi_1(X) > 0$ кажемо да је расподела позитивно асиметрична или асиметрична удесно, а за $\pi_1(X) < 0$ кажемо да је расподела негативно асиметрична или асиметрична улево.

Како коефицијент $\pi_1(X)$ не зависи од средње вредности, нити скалирања случајне промењиве, све расподеле истог типа имају исту асиметрију.

Спљоштеност расподеле се односи на брзину конвергенције ка нули крајева густине расподеле, као и концентрација око средње вредности. Користи се и термин дебљина репа расподеле и обично се упоређује са дебљином репа нормалне расподеле. Пирсон је дефинисао коефицијент спљоштености или киртозис као количник $\frac{\mu_4}{\mu_2^2}$. Како је киртозис нормалне расподеле једнак 3, често се као мера спљоштености узима вредност киртозиса умањена за 3.

Дефиниција 17 *Коефицијент спљоштености $\pi_2(X)$ је дефинисан са*

$$\pi_2(X) = \frac{\mu_4}{\mu_2^2} - 3.$$

За $\pi_2(X) > 0$ кажемо да је расподела са дебелим крајевима (реповима) или лептокиртична, за $\pi_2(X) < 0$ кажемо да је расподела са танким крајевима (реповима) или платикиртична, док за $\pi_2(X) = 0$ кажемо да расподела има крајеве (репове) исте дебљине као и нормална расподела или да је мезокиртична.

Спљоштеност је такође иста за све расподеле истог типа.

8 Ентропија

У теорији информација је важна количина информација садржана у случајном догађају, а као мера неизвесности случајне промењиве користи се појам ентропије који је увео Шенон[3].

[3] Claude Elwood Shannon (1916 -2001) - ??

Дефиниција 18 *За непрекидну случајну променљиву X са позитивном густином g ентропија $H(X)$ је дефинисана са*

$$H(X) = -\int_{-\infty}^{+\infty} g(x) \ln_2 g(x) dx.$$

Ако густина није свуда позитивна, онда се интеграл на десној страни претходне једнакости узима на скупу $\{x : g(x) > 0\}$.

Расподеле истог типа немају исту ентропију, али постоји једноставна релација за њихове ентропје.

Теорема 4 *Ако је $Y = a + bX$, где је $a \in R$ и $b > 0$, тада је*

$$H(Y) = H(X) + \ln b.$$

9 Карактеристична функција

Карактеристична функција непрекидне случајне променљиве је Фуријеова трансформација њене густине.

Дефиниција 19 *Нека је X непрекидна случајна променљива чија је густина g. Функција $\varphi : R \to C$, где је C скуп свих комплексних бројева, дефинисана са*

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} g(x) dx, \quad t \in R$$

је карактеристична функција за X.

Из дефиниције следи да свака непрекидна случајна променљива има карактеристичну функцију и да је

1. $\varphi(0) = 1$,
2. $\varphi(-t) = \overline{\varphi(t)}$,
3. $|\varphi(t)| \leq 1$.

Расподела случајне променљиве је јединствено одређена њеном карактеристичном фунцијом, при чему постоји и инверзна формула.

Теорема 5 *Ако је φ карактеристична функција случајне променљиве X и ако је*

$$\int_{-\infty}^{+\infty} |\varphi(t)| dt < \infty,$$

тада је густина g за X дата са

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt.$$

Осим тога, свака функција φ која је позитивно дефинитна и непрекидна на R и за коју је $\varphi(0) = 1$ је карактеристична функција неке случајне променљиве.

За карактеристичне функције променљивих са истим типом расподеле постоји веза.

Теорема 6 *Ако је $Y = a + bX$ и ако су φ_X и φ_Y карактеристичне функције за X и Y, тада је*

$$\varphi_Y(t) = e^{iat}\varphi_X(bt).$$

Помоћу карактеристичне функције могу се одредити моменти случајне променљиве.

Теорема 7 *Ако случајна променљива има моменат реда r и ако је њена карактеристична функција φ диференцијабилна n пута, , тада је*

$$m_r = \frac{1}{i^n}\varphi^{(n)}(0),$$

где је i имагинарна јединица ($i^2 = -1$).

Често се користи и следеће својство карактеристичне функције.

Теорема 8 *Ако су X и Y независне случајне променљиве са карактеристичним функцијама φ_X и φ_Y и ако је φ_{X+Y} карактеристична функција њиховог збира, тада је*

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t), \quad t \in R.$$

10 Генератриса момената

За случајне променљиве се дефинише и трансформација густине која је реална верзија Фуријеове трансформације.

Дефиниција 20 *Нека је X непрекидна случајна променљива чија је густина g. Функција $M : R \to R$ дефинисана са*

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} g(x)dx, \quad t \in R,$$

под условом да интеграл на десној стране ове једнакости постоји, је генератриса момената променљиве X.

На основу веза између функција φ и M,

$$\varphi(t) = M(it), \quad M(t) = \varphi(-it),$$

следи да се и помоћу генератрисе момената могу добити моменти за X. Наиме, ако за случајну променљиву X постоји моменат реда r, тада је

$$m_r = M^{(r)}(0).$$

Предност функција генератриса је у томе што су реалне функције, а предност карактеристичних функција је у томе што не постоји проблем конвергенције.

11 Оцене параметара

За оцену непознатог параметра неке расподеле на основу простог узорка $(X_1, \ldots, X_n)$ могу се користити различите статистике. У њима се често појављују стандардне статистике као што су:

1. узорачка средина $\overline{X}_n$ или $\overline{X}$ дефинисана са

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

2. узорачка дисперзија $\overline{S}_n^2$ или $\overline{S}^2$ дефинисана са

$$\overline{S}_n^2 = \frac{1}{n} \left(\sum_{i=1}^{n} X_i - \overline{X}_n \right)^2,$$

3. узорачка дисперзија $\overline{D}_n^2$ или $\overline{D}^2$ дефинисана са

$$\overline{D}_n^2 = \frac{1}{n} \left(\sum_{i=1}^{n} X_i - m \right)^2,$$

у случају да је познато математичко очекивање m посматране расподеле,

4. модификована узорачка дисперзија $\widehat{S}_n^2$ или $\widehat{S}^2$ дефинисана са

$$\widehat{S}_n^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i - \overline{X}_n \right)^2,$$

5. узорачки моменат $\overline{X}_n^r$ или $\overline{X}^r$ или M_r $(r \in N)$ дефинисан са

$$\overline{X}_n^r = \frac{1}{n} \sum_{i=1}^{n} X_i^r.$$

Наравно, важно је знати својства оцена.

Две основне технике за оцене параметара су метода момената и метода максималне веродостојности. Методу момената је увео Карл Пирсон 1984. године, а методу максималне веродостојности Фишер 1912. године. Метода момената је једноставнија, а метода максималне веродостојности даје асимптотски најефикасније оцене.

11.1 Својства оцена

Нека је $(X_1, \ldots, X_n)$ прост случајан узорак за обележје X и нека је θ непознати параметар у расподели тог обележја.

Дефиниција 21 *Статистика $\widehat{\theta}_n$ је непристрасна оцена за параметар θ ако је $E(\widehat{\theta}_n) = \theta$, а асимптотски непристрасна оцена ако $E(\widehat{\theta}_n) \to \theta$ када $n \to \infty$.*

Уместо термина *непристрасна* користе се и термини *центрирана* и *непомерена*. Оцена која није непристрасна зове се пристрасна. На пример, статистика $\overline{X}_n$ је непристрасна оцена за математичко очекивање, статистика D_n^2 је непристрасна оцена за варијансу, док је статистика S_n^2 пристрасна и асимптотски непристрасна оцена за варијансу.

Термине *пристрасна* и *непристрасна оцена* увео је Боули[4] 1897. године.

Следеће својство се односи на одступање реализоване вредности статистике од очекиване вредности.

Дефиниција 22 *Статистика $\widehat{\theta}_n$ је постојана оцена параметра θ ако $\widehat{\theta}_n$ конвергира у вероватноћи ка θ, односно ако за свако $\varepsilon > 0$ важи*

$$P\{|\widehat{\theta}_n - \theta| \geq \varepsilon\} \to 0 \ (n \to \infty).$$

Уместо термина *постојана* користе се и термини *конзистентна* и *стабилна*. Ако је оцена непристрасна и ако $D(\widehat{\theta}_n) \to 0$ када $n \to \infty$, тада је оцена $\widehat{\theta}_n$ стабилна. На пример, статистике $\overline{X}_n$, D_n^2 и S_n^2 су стабилне оцене.

Више различитих, непристрасних и стабилних оцена истог параметра, за исти обим узорка, могу се упоређивати помоћу својих дисперзија.

Дефиниција 23 *Оцена $\widehat{\theta}$ је ефикаснија од оцене $\widetilde{\theta}$ непознатог параметра θ ако је $D(\widehat{\theta}) < D(\widetilde{\theta})$.*

Доњу границу дисперзије непристрасних оцена даје Рао-Крамерова неједнакост. Оцена чија је дисперзија једнака тој доњој граници је *ефикасна оцена*. За анализу оцене није довољно знати само њено очекивање и дисперзију, већ треба одредити и њену расподелу.

11.2 Метода момената

Непрекидне расподеле су, под извесним условима, одређене својим моментима. На пример, ако је за низ момената (m_i) дате расподеле испуњен Карлеманов услов

$$\sum_{n=1}^{\infty} \frac{1}{m_{2n}^{1/2n}} = +\infty,$$

тада моменти једнозначно одређују ту расподелу. Самим тим, познавање момената је довољно за налажење непознатих параметара расподеле.

Ако су непознати параметри $\theta_1, \theta_2, \ldots, \theta_k$, довољно је узети k момената који зависе од тих параметара. Претпоставимо да су то првих k момента. У том случају имамо систем

$$m_r = f_r(\theta_1, \ldots, \theta_k), \quad r = 1, \ldots, k$$

који одређује непознате параметре. Међутим, како параметре оцењујемо на основу узорка, уместо момената узимамо узорачке моменте и онда из наведног система добијамо оцене непознатих параметара.

[4] Arthur Lyon Bowley (1869-1957) - енглески статистичар

Дефиниција 24 *Оцене параметара $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ по методи момената, на основу простог узорка $(X_1, \ldots, X_n)$ из неке расподеле, дате су системом*

$$\overline{X}_n^r = f_r(\widehat{\theta}_1, \ldots, \widehat{\theta}_k), \quad r = 1, \ldots, k.$$

Према томе,

$$\widehat{\theta}_i = \widehat{\theta}_i(\overline{X}_n, \overline{X}_n^2, \ldots, \overline{X}_n^k).$$

За реализовани узорак $(x_1, \ldots, x_n)$ добијају се конкретне оцене непознатих параметара.

Предност ове методе је једноставност, а недостатак је немогућност одређивања квалитета и интервала поверења за добијене оцене.

11.3 Метода максималне веродостојности

У овој методи оцена за непознати параметар θ се узима тако да вероватноћа реализације узорка $x_1, x_2, \ldots, x_n$ буде највећа. Претпоставимо да је $(X_1, \ldots, X_n)$ прост узорак из неке непрекидне расподеле чија густина g зависи од θ, при чему $\theta \in \Theta \subset R$. Функција L дефинисана са

$$L(x_1, \ldots, x_n; \theta) = \prod_{k=1}^{n} g(x_k; \theta)$$

представља густину расподеле узорка као случајног вектора. Обзиром да се реализоване вредности узорка концентришу у области у којој функција L има велике вредности, за оцену непознатог параметра се узима вредност која максимизира функцију L.

Дефиниција 25 *Оцена $\widehat{\theta}$ одређена са*

$$L(X_1, \ldots, X_n; \widehat{\theta}) = \max_{\theta} L(X_1, \ldots, X_n; \theta)$$

је оцена максималне веродостојности. Функција L је функција веродостојности.

Ако је Θ отворен скуп и ако је L диференцијабилна функција по θ, довољно је одредити стационарне тачке функције L и упоредити одговарајуће вредности. Пошто функције L и $\ln L$ имају исте тачке максимума, оцена се једноставније добија из једначине

$$\frac{\partial \ln L}{\partial \theta} = 0,$$

познате као *једначина веродостојности*. У неким случајевима она има само једно решење.

Теорема 9 *Ако постоји ефикасна оцена параметра θ, онда једначина веродостојности има јединствено решење.*

Метода максималне веродостојности може се применити и у случају да расподела има више непознатих параметара. Ако је $g(x;\theta_1,\ldots,\theta_k)$ густина расподеле која зависи од параметара $\theta_1,\ldots,\theta_k$, тада је функција веродостојности дата са

$$L(x_1,\ldots,x_n;\theta_1,\ldots,\theta_k)=\prod_{i=1}^{n} g(x_i;\theta_1,\ldots,\theta_k).$$

Ако функција L има парцијалне изводе по $\theta_1,\ldots,\theta_k$, оцена непознатих параметара добија се решавањем система

$$\frac{\partial}{\partial\theta_i}L(x_1,\ldots,x_n;\theta_1,\ldots,\theta_k)=0,\quad i=1,\ldots,k.$$

Овај систем је углавном нелинеаран и најчешће се решава нумеричким методама.

Оцене добијене методом максималне веродостојнсти могу бити и пристрасне и неефикасне. Међутим, при одређеним условима оцене су асимптотски ефикасне и имају асимптотски нормалну расподелу.

12 Трансформација случајне променљиве

Ако је позната функција расподеле F_X непрекидне случајне променљиве X и ако је f монотона функција, тада је $Y=f(X)$ такође непрекидна случајне променљива чија се функција расподеле једноставно одређује. Уз додатни услов диференцијабилности функције f лако се налази и густина случајне променљиве $f(X)$.

Теорема 10 *Нека је $Y=f(X)$ и нека је $f:R\to R$ строго монотона диференцијабилна функција. Ако је $h=f^{-1}$ (инверзна функција за f), тада је*

$$g_Y(y)=g_X(h(y))|h'(y)|,\quad y\in(a,b)$$

и $g_Y(y)=0$ за $x\notin(a,b)$, где је

$$a=\min\{f(-\infty),f(+\infty)\},\quad b=\max\{f(-\infty),f(+\infty)\}.$$

Густина случајне променљиве $f(X)$ се може одредити и ако је f део по део монотона. Нека је g строго монотона и диференцијабилна на интервалима $I_1,\ldots,I_k$ и нека је $f_j:I_j\to R$ за $j=1,\ldots k$ рестрикција функцијие f на интервалу I_j. Тада је

$$Y=f_1(X)+f_2(X)+\cdots+f_k(X),$$

па применом претходне теореме на случајне променљиве $f_1(X),\ldots,f_k(X)$ добијамо следеће тврђење.

Теорема 11 *Нека је $Y=f(X)$ и нека је $h_j=f_j^{-1}$ за $j=1,\ldots,k$. Тада је*

$$g_Y(y)=\sum_{j=1}^{k} f_X(h_j(y))|h_j'(y)|.$$

За налажење математичког очекивања случајне променљиве $Y = f(X)$ није неопходно знати густину g_Y.

Теорема 12 *Ако је* $Y = f(X)$, *тада је*

$$E(Y) = \int_{-\infty}^{+\infty} f(x)g(x)dx$$

под условом да наведени интеграл апсолутно конвергира.

12.1 Такијева трансформација

Амерички статистичар Таки[5] је 1962. године предложио трансформацију

$$Y = \begin{cases} \dfrac{X^\lambda - (1-X)^\lambda}{\lambda}, & \lambda \neq 0 \\ \ln \dfrac{X}{1-X}, & \lambda = 0 \end{cases}$$

где је X случајна променљива са стандардном униформном расподелом. За одређене вредности параметрта λ расподела добро апроксимира нормалну и студентову расподелу.

12.2 Бокс-Кокс трансформација

За позитивне случајне променљиве Бокс[6] и Кокс[7] су 1964. године увели трансформацију којом се добија приближно нормална расподела. Ако је X позтивна случајна променљива, тада случајна променљива Y дефинисана са

$$Y = \begin{cases} \frac{X^\lambda - 1}{\lambda,} & \lambda \neq 0 \\ \ln X, & \lambda = 0 \end{cases}$$

за погодно изабрану вредност константе λ има приближно нормалну расподелу.

12.3 Џонсонове трансформације

Енглески статистичар Џонсон[8] је 1949. године предложио трансформације

$$X = a + b \ln \frac{Y}{1-Y},$$

$$X = a + b\, arcsh\, Z,$$

где је X случајна променљива са стандардном нормалном расподелом.

Расподела случајне променљиве Y је позната као Џонсонова S_B расподела, а расподела случајне променљиве Z као Џонсонова S_U расподела.

---

[5] John Wilder Tukey (1915-2000) - aмерички стаистичар

[6] George Box (1919-) - енглески статистичар

[7] David Roxbee Cox (1924-) - енглески статистичар

[8] Norman Lloyd Johnson (1917-) - енглески статистичар

13 Моделирање случајних променљивих

Моделирање непрекидне случајне променљиве зависи од карактеристика њене расподеле. У принципу постоје опште методе, као што су *метода инверзне функције*, *Нојманова метода* и друге, али постоје и специфични поступци моделирања за поједине расподеле.

13.1 Метода инверзне функције

Ова метода заснива се на једној једноставној чињеници. Ако је X случајна променљива с непрекидном и монотоно растућом функцијом расподеле F, случајна променљива $Y = F(X)$ има униформну расподелу на $(0, 1)$. Према томе, функција расподеле случајне променљиве $F^{-1}(Y)$ је F, односно $X = F^{-1}(Y)$.

За генерисање вредности x случајне променљиве X довољно је генерисати вредност y из униформне расподеле $U(0, 1)$ и узети да је $x = F^{-1}(y)$. На тај начин се проблем своди на генерисање вредности случајне променљиве са униформном расподелом, а за то постоје већ готове рутине у свим статистичким програмским пакетима. То су, заправо, *случајни бројеви*.

Генерисање расподеле методом инверзне функције је, дакле, врло једноставно, али је потребно да знамо аналитички облик инверзне функције. Међутим, у многим случајевима то није могуће добити. Због тога ова метода не може да се примени за генерисање, на пример нормалне, бета, гама или хи квадрат расподеле.

Метода се може проширити на случај када је

$$g(x) = \sum_{k=1}^{n} p_k g_k(x),$$

где је $p_1 + \cdots + p_n = 1$ и где густинама g_k одговарају функције расподеле F_k $(k = 1, \ldots, n)$ које имају аналитички израз за инверзну функцију. Алгоритам је тада следећи:

1. Изабрати индекс k са вероватноћом p_k.
2. Генерисати случајну вредност y из униформне расподеле $U(0, 1)$.
3. Узети да је $x = F_k^{-1}(y)$.

13.2 Нојманове методе

Нојманове методе или *методе прихватања и одбацивања* користе се за генерисање расподеле са густином g која је позитивна на интервалу $[x_{\min}, x_{\max}]$. Углавном се у њима користи алгоритам у којем се проверава одређени услов и зависно од испуњења тог услова одређена вредност се прихвата или одбацује као случајан број из расподеле са густином g.

Једна од тих метода захтева познавање случајне променљиве Y коју знамо да генеришемо и чија расподела има густину h за коју важи

$$h(x) \le c g(x), \quad x \in [x_{\min}, x_{\max}]$$

за неки реалан број $c > 0$. Метода може да се опише следећим алгоритмом:

1. Нека је x генерисана вредност случајне променљиве Y.
2. Генерисати случајан број u из униформне расподеле $U(0,1)$.
3. Ако је $u < \dfrac{g(x)}{ch(x)}$, тада је x тражена вредност случајне променљиве X са густином g. У противном, поступак се понавља (од 1. корака).

Ова метода се заснива на чињеници да је

$$P\left(Y \leq x \,\middle|\, U \leq \frac{g(Y)}{ch(Y)}\right) = \int_{-\infty}^{x} g(y)dy,$$

где је U случајна променљива са униформном расподелом $U(0,1)$. Остале Нојманове методе су мање или веће модификације ове идеје.

14 Тестови сагласности

За тестирање сагласности расподеле посматраног обележја са претпостављеном расподелом постоји више различитих тестова. Тестови се могу поделити у две групе према томе да ли расподела тест статистике зависи од претпостављене расподеле или не. У тестове слободне од расподеле спадају, на пример, Пирсонов χ^2 тест, Колмогоровљев[9] тест, Куиперов[10] тест, фон Мизисов[11] тест. У другу групу спадају, на пример, Андерсон[12]-Дарлингов[13] тест, разне модификације Колмогоровљевог теста уколико параметри претпостављене расподеле нису познати, Шапиро[14]-Вилков[15] тест. За тестирање прилагођености података нормалној расподели постоји још и специфични тестови засновани на поређењу емпиријских коефицијената асиметрије и спљоштености са одговарајућим коефицијентима нормалне расподеле.

14.1 Пирсонов тест

Најпопуларнији је χ^2-тест који је увео Пирсон 1900. године. Тест има облик

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

где су O_i опсервиране, а E_i очекиване фреквенције (под претпоставком одређене расподеле) у интервалима I_i $(i = 1, \ldots, k)$, таквим да је $\cup_{i=1}^{k} I_i = R$

[9] Андрей Николаевич Колмогоров (1903-1987) - руски математичар
[10] ??? (??-??) - ??????
[11] Richard Martin Edler von Mises (1883-1953) - аустријски инжењер и математичар
[12] Theodore Wilbur Anderson (1918-) - амерички статистичар
[13] Davis Darling (??-??) - ?????
[14] Samuel Shapiro (1930-) - амерички статистичар
[15] Martin Bradbury Wilk (1922-) - амерички математичар

и $I_i \cap I_j = \emptyset$ за $i \neq j$. Претпостављајући нулту хипотезу да су опсервиране и очекиване фреквенције статистички значајно не разликују, тест статистика има расподелу $\chi^2(k-p-1)$, где је p број оцењених параметара на основу k опсервираних вредности.

14.2 Тестови којима се пореде емпиријска и теоријска функција расподеле

На основу разлике између емпиријске функције расподеле F_n (простог случајног узорка $X_1, \dots, X_n$) и теоријске функције расподеле F треба тестирати нулту хипотезу да обележје X има расподелеу F. У тест статистикама се узимају неке функције разлика $F_n(x) - F(x)$ за $x \in R$. На пример, тест статистика Колмогоровљевог теста је

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

тест статистика Куиперовог теста је

$$V_n = \sup_x |F_n(x) - F(x)| - \inf_x |F_n(x) - F(x)|,$$

док је тест статистика фон Мизисовог теста

$$W_n = \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 dF(x).$$

15 Графичке методе испитивања сагласности

Поред тестова за испитивање сагласности обележја са претпостављеном расподелом користе се и графичке методе као што су *дијаграм квантила* ($Q - Q$ дијаграм) и *дијаграм вероватноћа* ($P - P$ дијаграм) које се базирају на статистици поретка $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

$Q - Q$ дијаграм је скуп тачака чије су координате $\left(X_{(i)}, F^{-1}(p_i)\right)$, где се за p_i узима нека од вредности $\dfrac{i-1/2}{n}, \dfrac{i}{n+1}, \dfrac{i-3/8}{n+1/4}$, док је $P-P$ дијаграм скуп тачака $\left(p_i, F(X_{(i)})\right)$.

У случају да су подаци сагласни са претпостављеном расподелом, тачке $Q - Q$ дијаграма приближно припадају правој. Исто важи и за $P - P$ дијаграм.