7. МОДЕЛИ РАСПОДЕЛА СЛУЧАЈНИХ ПРОМЕНЉИВИХ

На основу природе појаве коју анализирамо, често можемо претпоставити да расподела случајне променљиве *X* припада једној класи расподела, одређеној са једним или више непознатих параметара. Ако је та класа изражена функционалном везом

$$P(x;\theta_1,\theta_2,...\theta_k) \tag{7.1}$$

онда кажемо да функција (7.1) представља модел расподеле случајне променљиве *X*.

Модел расподеле (7.1) садржи све информације о посматраној појави која се изражава преко случајне променљиве *X*.

Постоје модели расподела који се најчешће користе и неки од њих ће бити детаљније разматрани у поглављима која следе. Биће наведени најчешће коришћени модели прекидних и непрекидних расподела случајних променљивих.

7.1. МОДЕЛИ ПРЕКИДНИХ РАСПОДЕЛА

Биномна расподела

Дефиниција 7.2: *Случајна променљива X има Биномну расподелу, ако може узети једну вредност k из низа ненегативних целих бројева*

$$\{0,1,2,\ldots,n\}$$

са вероватноћом:

$$P_n(k)=P(X=k)=\binom{n}{k}p^k q^{n-k} \tag{7.2}$$

при чему је $0<p<1$, $p+q=1$, а $\binom{n}{k}=\dfrac{n!}{(n-k)!\,k!}$ представља биномни коефицијент.

Чињеницу да случајна променљива *X* има Биномну расподелу означаваћемо са:

$$X : \mathrm{B}\,(n;p).$$

Закон вероватноћа Биномне расподеле дат је следећом табелом:

X	0	1	2	...	n
p(x)	$\binom{n}{0}p^0q^{n-0}$	$\binom{n}{1}pq^{n-1}$	$\binom{n}{2}p^2q^{n-2}$	...	$\binom{n}{n}p^nq^0$

Бројеви n і p су параметри модела.

На основу биномног обрасца, лако је проверити да је збир биномних вероватноћа једнак јединици. Заиста,

$$\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}=(p+q)^n=1$$

Функција генератриса Биномне расподеле је:

$$g(t)=\sum_{k=0}^{n}e^{kt}\binom{n}{k}p^k q^{n-k}=\sum_{k=0}^{n}\binom{n}{k}\left(pe^t\right)^k q^{n-k}=\left(\mathrm{q}+\mathrm{pe}^{\mathrm{t}}\right)^n \tag{7.3}$$

Извод функције $g(t)$ у тачки $t=0$, представља очекивану вредност:

$$E(X)=np \tag{7.4}$$

Преко другог извода функције $g(t)$ у тачки $t=0$ добија се други обични моменат, па се лако израчунава вредност варијансе:

$$\sigma^2=npq \tag{7.5}$$

Може се проверити да је коефицијент асиметрије:

$$\beta_1 = \frac{q-p}{\sqrt{npq}} \tag{7.6}$$

а коефицијент спљоштености:

$$\beta_2 = 3 + \frac{1-6pq}{npq} \tag{7.7}$$

Из (7.6) и (7.7), следи да Биномна расподела тежи симетричној расподели кад $n \to \infty$. Исто тако, она тежи расподели са „нормалном” спљоштеношћу.

Специјално, кад је $p = q = 0{,}5$, Биномна расподела је симетрична расподела за све вредности *n*.

Биномна расподела се користи у ***Бернулијевој (Bernulli) шеми*** независних понављања експеримента, обрађеној у поглављу 4.2. Одиграла је изузетно важну улогу у теорији вероватноће, њеној примени и у развоју статистике. Вероватноће случајних догађаја везаних за Бернулијеву шему експеримената могу се одредити у оквиру Биномне расподеле.

Пуасонова расподела

Пуасонова расподела названа је по француском математичару чије име је Simeon D. Poisson (1781-1840). Она описује вероватноћу да ће случајни догађај наступити у одређеном временском или просторном интервалу под одређеним условима, при чему је вероватноћа догађаја веома мала, а број покушаја (временски или просторни интервал) веома велики. Посматрани догађај се, дакле, ретко реализује. Пуасоновом расподелом се описује, на пример, број грешака по страници књиге, број жртава аутомобилских удеса у току једног месеца у неком граду и слично.

Дефиниција 7.3: *Случајна променљива X има Пуасонову расподелу ако може узети вредност k из низа ненегативних целих бројева*

$$\{0, 1, 2, 3, \ldots\}$$

и то са вероватноћом

$$P(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

при чему је $\lambda > 0$, реални број и представља параметар расподеле.

Чињеницу да случајна променљива има Пуасонову расподелу са параметром λ означавамо са:

$$X : \mathcal{P}(\lambda)$$

Закон вероватноћа дат је табелом:

X	0	1	2	...	k	...
p(x)	$\frac{\lambda^0}{0!}e^{-\lambda}$	$\frac{\lambda^1}{1!}e^{-\lambda}$	$\frac{\lambda^2}{2!}e^{-\lambda}$	...	$\frac{\lambda^k}{k!}e^{-\lambda}$	...

Зна се да је

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Зато је збир вероватноћа код Пуасонове расподеле једнак:

$$\sum_{k=0}^{\infty} P(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

Очекивана вредност случајне променљиве која има Пуасонову расподелу је:

$$m = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \qquad (7.9)$$

Други обичпi моменат је једнак:

$$m_2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} [k(k-1) + k] \frac{\lambda^k}{k!} =$$

$$= e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda$$

Варијанса случајне променљиве X је, према томе:

$$\sigma^2 = m_2 - m^2 = \lambda \qquad (7.10)$$

$$\mu_3 = \lambda$$

$$\mu_4 = 3\lambda^2 + \lambda$$

па је коефицијент асиметрије једнак:

$$\beta_1 = \frac{1}{\lambda}$$

а коефицијент спљоштености

$$\beta_2 = 3 + \frac{1}{\lambda}$$

Геометријска расподела

Посматрајмо статистички експеримент чији су могући исходи случаијни догађај A, чија је вероватноћа $P(A) = p$, и догађај $\overline{A}$ са вероватноћом $P(\overline{A}) = 1 - p = q$. Нека је у сваком (независном) понављању експеримента вероватноћа реализације догађаја A иста и једнака p. Док се у Бернулијевој шеми тражила вероватноћа да се у n понављања експеримента догађај A реализује тачно k пута, сада се посматра случајна променљива X која представља број понављања описаног експеримента до прве реализације догађаја A. Овако дефинисана случајна променљива X може да узме вредности 1, 2, 3, ... Вероватноћа да случајна променљива X узме вредност x дата је изразом:

$$p(x) = P(X = x) = q^{x-1} p \qquad \text{за } x = 1, 2, 3 \ldots \tag{7.15}$$

и представља вероватноћу да се у првих $x - 1$ понављања експеримента није реализовао посматрани догађај A, а да се реализовао у x-том понављању.

Формулом (7.15) дефинисан је закон вероватноћа Геометријске расподеле. Овај назив расподеле проистекао је из чињенице што вероватноће ове расподеле чине геометријски низ. Коришћењем обрасца за збир бесконачног опадајућег геометријског низа показује се да је збир свих вероватноћа једнак јединици:

$$\sum_{x=1}^{\infty} p(x) = p \sum_{x=1}^{\infty} q^{x-1} = p \frac{1}{1-q} = \frac{p}{p} = 1$$

Функција генератриса Геометријске расподеле је:

$$g(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} p q^{x-1} = \frac{pe^t}{1 - qe^t}$$

Математичко очекивање и варијанса су:

$$m = \frac{1}{p}; \qquad \sigma^2 = \frac{q}{p^2}$$

Вероватноћа да се до прве реализације догађаја A експеримент понови највише k пута износи:

$$p(X \le k) = p_1 + p_2 + \ldots + p_k = p \sum_{x=1}^{k} q^{x-1} = 1 - q^k$$

Овај резултат може се добити и нешто другачије: ако се експеримент понавља више од k пута, то значи да се догађај A није реализовао у првих k понављања експеримента. Овом случају одговара вероватноћа q^k, па је тада вероватноћа супротног догађаја једнака $1 - q^k$.

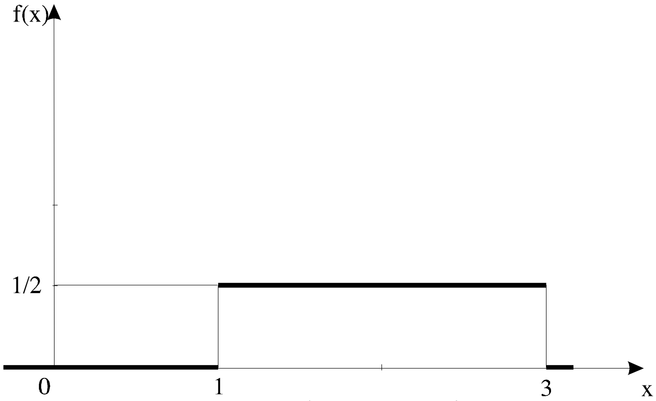
7.2. МОДЕЛИ НЕПРЕКИДНИХ РАСПОДЕЛА

Униформна расподела

Дефиниција 7.17: *Случајна промењљива X има* ***Униформну расподелу*** *ако је њен закон вероватноћа дат функцијом:*

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha}, & \alpha<x<\beta\\ 0, & \text{за остале } x\end{cases}$$

Закон вероватноћа (функција густине) за случајну промењљиву која има Униформну расподелу са параметрима $\alpha=1$ и $\beta=3$ дат је на слици:

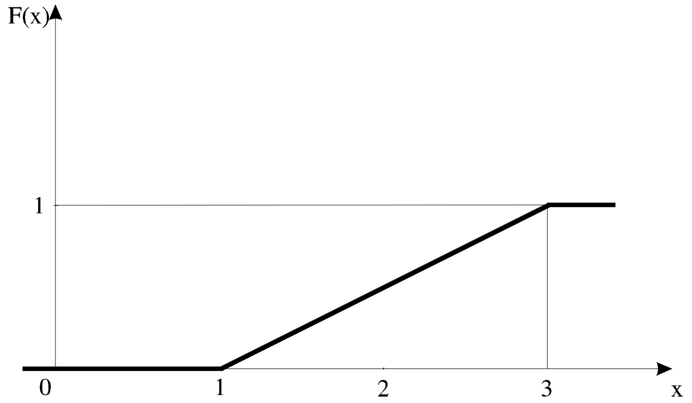


Funkcija gustine sl. pr. X

Функција расподеле за случајну промењљиву X која има Униформну расподелу са параметрима α и β је:

$$F(x)=P\{X\leq x\}=\begin{cases}0 & ; \quad x<\alpha\\ \dfrac{x-\alpha}{\beta-\alpha} & ; \quad \alpha\leq x\leq\beta\\ 1 & ; \quad x>\beta\end{cases}$$

За $\alpha=1$ и $\beta=3$, график ове функције је приказан на слици:



Funkcija raspodele sl. pr. X

Карактеристике ове расподеле су следеће:

Математичко очекивање:

$$E(X) = \frac{\alpha + \beta}{2}$$

Варијанса:

$$\sigma^2 = \frac{(\beta - \alpha)^2}{12}$$

Оса симетрије закона вероватноћа случајне промењљиве која има Униформну расподелу са параметрима α и β је:

$$x = \frac{\alpha + \beta}{2}$$

Нормална расподела

Приликом решавања проблема граничне функције Биномне расподеле, De Moivre је 1733. године открио Нормалну расподелу, што је остало доста незапажено све до радова C.F.Gauss-а 1809. године и P.S.Laplace-а 1812. године. Ова два аутора су развили теорију ***грешака опсервација*** и дошли до Нормалне расподеле. Laplace је доказао тзв. ***централну граничну теорему*** и дао одговоре на различита питања из теорије вероватноће.

Под њиховим утицајем се дуго сматрало да се статистичке расподеле, практично, свих врста података приближавају Нормалној расподели као идеалном облику и да је само потребно обезбедити довољно велики број тачних опсервација. Одступања сваке случајне варијабле од њене очекиване вредности посматране су као „грешке” и, на основу „закона грешака”, описане су Нормалном расподелом. Показало се да је то јесте тачно у многим физичким, астрономским, демографским и биолошким мерењима, а такође и у техничким и економским наукама. Ова чињеница омогућава да се, у оквиру статистичког закључивања и доношења одлука базираних на статистичком закључивању, велики део метода базира на моделима који су генерисани Нормалном расподелом.

<u>Дефиниција 7.5</u>: *Случајна промењива X има* ***Нормалну расподелу*** *ако је њен закон вероватноћа дат функцијом*

$$f(x)=\frac{1}{b\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2b^2}} \text{ за } x\in(-\infty,\infty) \tag{7.20}$$

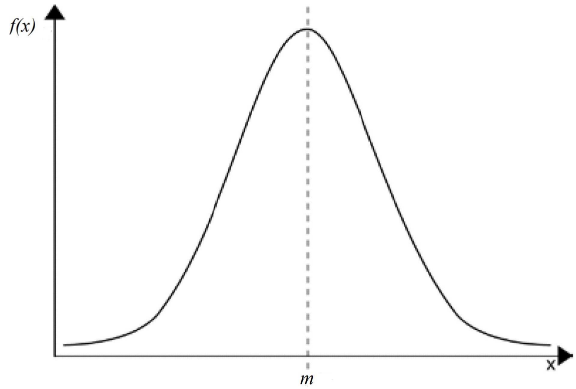
Вредности a и b^2 су параметри модела, a је било који реални број, а b^2 је позитивни реални број. Нормалну расподелу означаваћемо са:

$$X : N(a;b^2)$$

Функција (7.20) је симетрична у односу на праву $x=a$. Има максимум у тачки $x=a$. Максимална вредност функције је једнака:

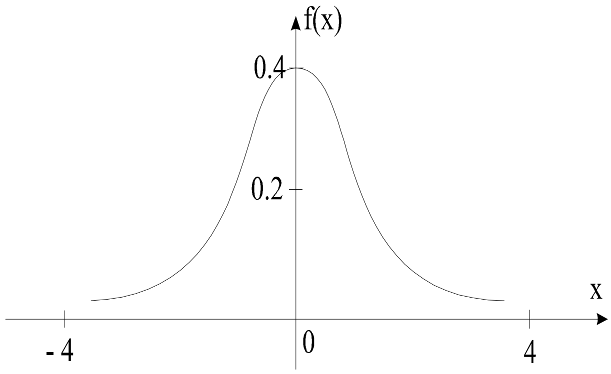
$$f(a)=\frac{1}{b\sqrt{2\pi}}$$

Графички приказ функције (7.20), дат је на слици:



Слика 7.1. Нормална расподела

Графички приказ функције (7.20), за $a=0$, дат је на слици:



Слика 7.2. Стандардизована Нормална расподела

Очекивана вредност случајне промењиве X која има Нормалну расподелу одређује се преко интеграла:

$$m=E(X)=\frac{1}{b\sqrt{2\pi}}\int_{-\infty}^{\infty}xe^{-\frac{(x-a)^2}{2b^2}}dx$$

Увођењем смене $y=\dfrac{x-a}{b}$ и парцијалном интеграцијом, добија се да је

$$m=a \tag{7.21}$$

Варијансу случајне промењљиве X са Нормалном расподелом израчунава се преко интеграла:

$$\sigma^2=E(X-m)^2=\frac{1}{b\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-a)^2 e^{-\frac{(x-a)^2}{2b^2}}dx$$

Увођењем смене $y=\dfrac{x-a}{b}$, и парцијалном интеграцијом, добија се да је

$$\sigma^2=b^2 \tag{7.22}$$

Из (7.21) и (7.22) види се да су a и b^2, очекивана вредност и варијанса случајне промењљиве X. Зато се најчешће за цлучајну промењљиву са Нормалном расподелом и користи ознака:

$$X:N\left(m;\sigma^2\right)$$

а њен закон вероватноћа дат је у облику:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Специјално, када је $m=0$ *и* $\sigma^2=1$*, случајна промењљива са законом вероватноћа датим тзв. Gauss-овом кривом:*

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

има ***Стандардизовану нормалну расподелу*** *чије вредности су дате у табелама или у статистичким пакетима програма.*

Вредности функције расподеле $F(x)$, за стандардизовану Нормалну расподелу, дате су у табелама и означене су са $\Phi(x)$.

За одређивање вероватноће да ће се вредности случајне промењљиве X, која има Нормалну расподелу са параметрима m и σ^2, наћи у интервалу $\left(x_1;x_2\right)$ најчешће се користе вредности функције стандардизоване Нормалне расподеле дате у табелама. Наиме, из једначине

$$P\left\{x_1<X\le x_2\right\}=\frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

после смене

$$y=\frac{x-m}{\sigma}$$

добиће се да је

$$P\left\{x_1 < X \le x_2\right\} = \frac{1}{\sqrt{2\pi}} \int_{\frac{x_1-m}{\sigma}}^{\frac{x_2-m}{\sigma}} e^{-\frac{x^2}{2}} dx = \Phi\left(\frac{x_2-m}{\sigma}\right) - \Phi\left(\frac{x_1-m}{\sigma}\right)$$

На основу наведеног се може закључити следеће: ако важи

$X:\ N(m,\sigma^2)$, тада $Y = \dfrac{X-m}{\sigma}\ :\ N(0,1)$. Наведена трансформација се често назива *стандардизацијом* случајне променљиве X.

Функција генератриса Нормалне расподеле је функција

$$g(t) = e^{tm+\frac{t^2\sigma^2}{2}}$$

Израчунавањем обичних момената преко функције $g(t)$ и одговарјућих централних момената, лако је проверити да је

$$\beta_1 = 0$$
$$\beta_2 = 3$$

из чега се закључује да је Нормална расподела симетрична и да има нормалну спљоштеност.

Хи-квадрат расподела (χ^2)

Хи-квадрат расподела дефинише се као расподела суме квадрата независних случајних променљивих које имају Нормалну расподелу, са математичким очекивањем 0 и стандрадном девијацијом 1. Ова расподела је важна у статистици јер описује расподелу вероватноћа које се користе у неколико најчешће коришћених статистичких процедура.

Дефиниција 7.7: *Случајна променљива X има Хи-квадрат расподелу са* n *степени слободе ако је њен закон вероватноћа дат функцијом*

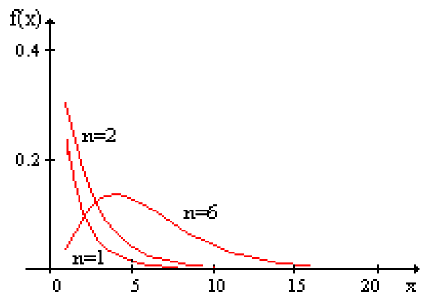
$$k_n(x) = \begin{cases} c_n x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{за x} > 0 \\ 0, & \text{за x} \le 0 \end{cases}$$

при чему је n *природан број, а константа* c_n *је дата са*

$$c_n = \frac{1}{\int_0^\infty x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx} = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)},$$

где је $\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \quad p > 0$ *је Гама функција или Ојлерова функција друге врсте.*

Закони вероватноћа за *Хи*-квадрат расподелу и за неке вредности *n*, приказани су на слици:



Слика 7.3. Хи - квадрат расподеле

Случајну променљиву која има *Хи*-квадрат расподелу са *n* степени слободе означаваћемо са

$$X : \chi_n^2 .$$

Функција расподеле ове случајне променљиве означава се са

$$K_n(z) = C_n \int_0^z x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx$$

Њене вредности су дате табелама за степене слободе од 1 до 30. За веће степене слободе могу се користити приближне вредности из табеле функције Нормалне расподеле.

Очекивана вредност случајне променљиве која има *Хи*-квадрат расподелу са *n* степени слободе је:

$$E(x) = n$$

а варијанса:

$$\sigma^2 = E(x-m)^2 = 2n$$

Обичан моменат реда *k* дат је изразом

$$m_k = n(n+2)(n+4)\ldots(n+2k-2)$$

тако да је

$$\mu_3 = 8n$$
$$\mu_4 = 12n(n+4)$$

Коефицијент асиметрије има вредност

$$\beta_1 = \sqrt{\frac{8}{n}}$$

а коефицијент спљоштености је

$$\beta_2 = 3 + \frac{12}{n}$$

Из формуле за коефицијент асиметрије се види да *Хи*-квадрат расподела тежи симетричној расподели кад $n \to \infty$, а из формуле за коефицијент спљоштености се види да њена спљоштеност тежи нормалној.

Студентова расподела

Студентову расподелу дефинисао је W.S. Gosset у раду објављеном под псеудонимом „Студент”.

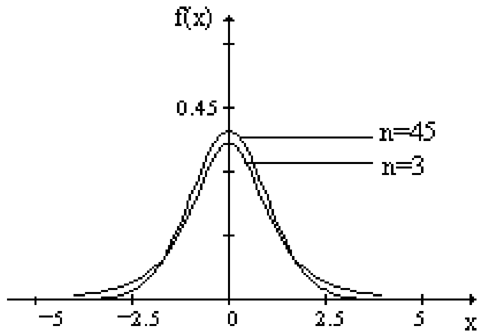
Дефиниција 7.8: *Случајна променљива X има* ***Студентову расподелу са n степени слободе*** *ако је њен закон вероватноћа дат функцијом*

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi \cdot n} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < x < \infty$$

Случајну променљиву која има Студентову расподелу са n степени слободе означаваћемо са

$$X: t_n$$

На слици су приказане функције $f_n(x)$, за $n = 3$ и $n = 45$:



Слика 7.4. Студентова расподела

Закон вероватноћа Студентове расподеле дат је симетричном функцијом са максималном вредношћу за $x = 0$. Очекивана вредност случајне променљиве која има Студентову расподелу је:

$$m = E(x) = 0$$

Обични и централни моменти су међусобно једнаки. Варијанса је једнака:

$$\sigma^2 = \frac{n}{n-2}, \qquad n>2$$

коефицијент асиметрије $\beta_1 = 0$, а коефицијент спљоштености је једнак

$$\beta_2 = 3 + \frac{6}{n-4}, \quad n>4$$

Вредности функције Студентове расподеле налазе се у табелама, а када $n \to \infty$, Студентова расподела тежи Стандардизованој нормалној расподели.

За $X: N(0,1)$ и $Y:\chi^2_n$, које су међусобно независне, важи:

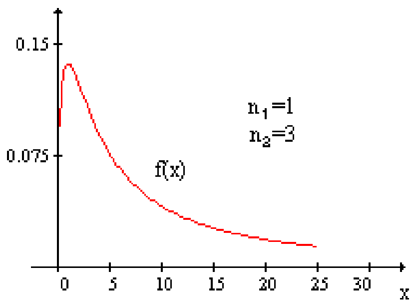
$$Z=\frac{X}{\sqrt{\frac{Y}{n}}}:\ t_n\,.$$

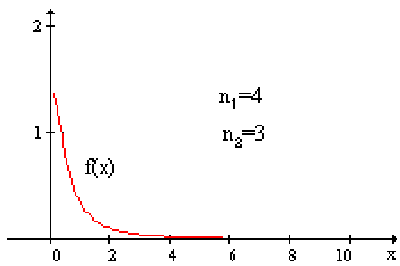
F-расподела

<u>Дефиниција7.10</u>: *Случајна промењљива X има* ***F-расподелу са n_1 и n_2 степени слободе****, ако је њен закон вероватноћа дат функцијом*

$$f_{n_1,n_2}(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\cdot x^{\frac{n_1}{2}-1}\cdot\left(1+\dfrac{n_1x}{n_2}\right)^{-\frac{n_1+n_2}{2}} & \text{за } x>0\\ 0 & \text{за } x\le 0\end{cases}$$

Графици закона вероватноћа *F*-расподеле за различите вредности параметара приказани су на следећим сликама:





Слика 7.5. F-расподела

Очекивана вредност *F*-расподеле једнака је

$$E(X)=\frac{n_2}{n_2-2}$$

а варијанса је

$$\sigma^2 = \frac{2n_2^2\left(n_1 + n_2 - 2\right)}{n_1\left(n_2 - 2\right)^2\left(n_2 - 4\right)}$$

При одређивању вероватноћа случајне променљиве са *F*-расподелом користимо табелу, која садржи вредности F_0 за које је функција расподеле једнака $1-\alpha = 0,95$ и $1-\alpha = 0,99$, а за различите степене слободе.

Ако је случајна променљива X дефинисана као количник

$$X = \frac{\dfrac{Y}{n_1}}{\dfrac{Z}{n_2}} = \frac{n_2 Y}{n_1 Z},$$

при чему су Y и Z независне случајне променљиве које имају Хи-квадрат расподеле са n_1 *и* n_2 *степени слободе, респективно. Тада X има* ***F-расподелу са*** $\boldsymbol{n_1}$ ***и*** $\boldsymbol{n_2}$ ***степени слободе***.