EMPIRIJSKE VJEROJATNOSTI

empirijska vjerojatnost za niz u opadanju: $\mathbf{F(X \geq x_m) = m/N}$ (a)

empirijska vjerojatnost za niz u porastu: $\mathbf{F(X \leq x_m) = (N-m+1)/N}$

obzirom da je: $F(X \geq x_m) + F(X \leq x_m) = 1$

vrijedi: $\mathbf{F(X \geq x_m)} = 1 - F(X \leq x_m) = 1 - (N-m+1)/N \mathbf{= (m-1)/N}$ (b)

usporedba (a) i (b) navodi na zaključak da je: $\mathbf{(m-1)/N} < \mathbf{F(X \geq x_m)} < \mathbf{m/N}$

Zbog toga su mnogi autori predložili kompromisne izraze za empirijsku vjerojatnost.

primjer iz tablice: $F(X \geq \mathbf{a_3}) = \mathbf{m/N} = 3/6 = 1/2$ (a)

$F(X \geq \mathbf{a_3}) = \mathbf{(m-1)/N} = (3-1)/6 = 2/6 = 1/3$ (b)

zaključak: $\mathbf{1/3} < \mathbf{F(X \geq x_m)} < \mathbf{1/2}$

Redni broj	1	2	3	4	5	6
niz u opadanju	$a_1 >$	$a_2 >$	$\mathbf{a_3} >$	$a_4 >$	$a_5 >$	a_6
niz u porastu	$a_6 <$	$a_5 <$	$a_4 <$	$\mathbf{a_3} <$	$a_2 <$	a_1
			m=3	N-m+1		N=6

KOMPROMISNI IZRAZI ZA EMPIRIJSKU VJEROJATNOĆU

Hazen: $F(X \geq x) = \dfrac{2m-1}{2N}$	Blom: $F(X \geq x) = \dfrac{m-3/8}{N+1/4}$
Weibull: $F(X \geq x) = \dfrac{m}{N+1}$	Tukey: $F(X \geq x) = \dfrac{3m-1}{3N+1}$
Čegodajev: $F(X \geq x) = \dfrac{m-0,3}{N+0,4}$	Gringorten: $F(X \geq x) = \dfrac{m-0,44}{N+0,12}$

PARAMETRI EMPIRIJSKIH RASPODJELA VJEROJATNOSTI

Tablični pregled učestalosti *(frekvencija)* i odgovarajući grafički prikazi *(krivulje trajanja i učestalosti)* predstavljaju jedan vid organiziranja serija podataka hidroloških mjerenja. Njihove numeričke karakteristike interpretiraju se slijedećim vrstama parametara:

- **Parametri koji karakteriziraju centralnu tendenciju i lokaciju raspodjele učestalosi:** *srednja vrijednost, geometrijska sredina, harmonijska sredina, medijana, mod;*

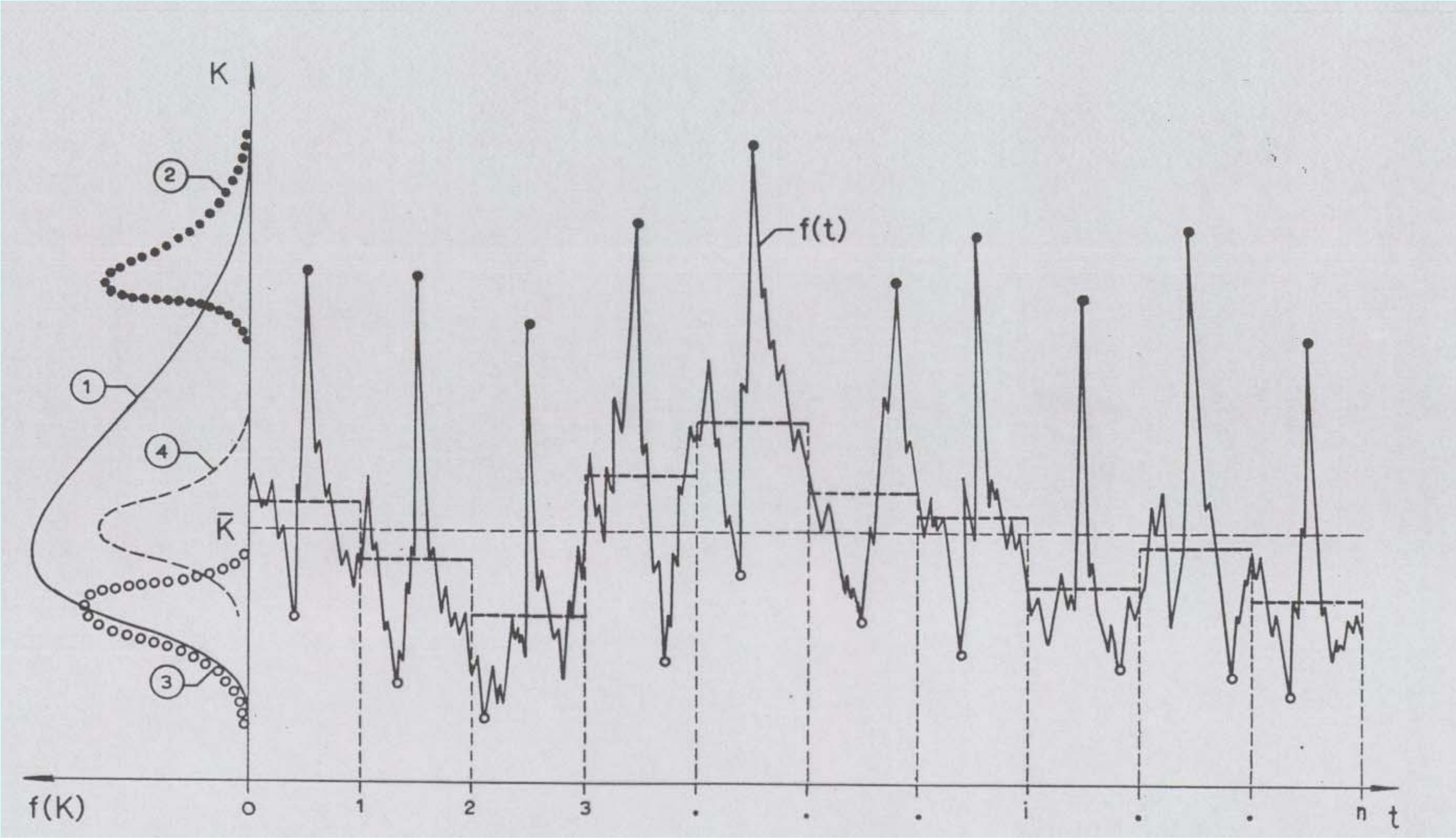
- **Parametri koji kvantificiraju** *(pokazuju)* **disperziju vrijednosti slučajne varijable oko srednje vrijednosti:**

  interval varijacije, varijanca, standardna devijacija, koeficijent varijacije i dr.

- **Parametri s kojima se precizira *(određuje)* asimetrija i spljoštenost krivulje učestalosti:**

  koeficijent asimetrije, mjera spljoštenosti.

Najčešće statističke obrade hidroloških veličina



Vremenska serija hidrološke varijable s krivuljama učestalosti: (1) svih članova uzorka, (2) maksimuma, (3) minimuma, (4) srednjih vrijednosti specificiranih vremenskih intervala

Najčešće vjerojatnosne analize u hidrologiji

Temeljem vremenskih serija hidroloških podataka najćešće se provode analize ili proračuni vjerojatnosti pojavljivanja slijedećh hidrolških veličina:

- maksimalni vodostaji
- minimalni vodostaji
- srednji godišnji protoci
- srednji godišnji volumeni protekle vode
- maksimalni protoci
- minimalni protoci
- maksimalni volumeni velikih vodnih valova
- srednji godišnji pronos riječnog nanosa
- godišnji volumeni pronosa riječnog nanosa
- maksimalni, minimalni i srednji nivoi podzemnih voda

Cilj vjerojatnosnih analiza je pridružiti bilo kojem promatranom mjernom iznosu hidrološke veličine (tj. slučajne varijable) vjerojatnost pojavljivanja upravo tog promatranog mjernog iznosa

Formiranje skupova podataka za vjerojatnosne analize u hidrologiji

Za vjerojatnosne analize u hidrologiji **biraju se** iz ukupnog fonda hidroloških podataka oni **podaci koji su reprezentativni za problem koji se analizom treba riješiti:**

Za **analize srednjih voda** biraju se srednje vrijednosti iz specificiranih vremenskih intervala *(npr. srednji godišnji vodostaji ili protoci).*

Za **analize velikih voda** biraju se podaci samo iz domene velikih voda. U praksi se primjenjuju dva osnovna principa formiranja reprezentativnog niza podataka o velikim vodama i to:

niz maksimuma - *iz svake godine (tj. specificiranog vremenskog intervala) bira se samo maksimalan podatak (vodostaj, protok ili volumen vala velike vode – zavisno o problem koji se reješava),*

niz prekoračenja praga (*treshhold princip*) - *iz svake godine (tj. specificiranog vremenskog intervala) biraju se svi oni podaci koji su veći od neke odabrane vrijednosti (tj. praga vodostaja, protoka ili volumena vala velike vode – zavisno o tome koji se problem reješava),*

Za **analize malih voda** biraju se podaci samo iz domene malih voda. U praksi se također primjenjuju dva osnovna principa formiranja reprezentativnog niza podataka:

niz minimuma - *iz svake godine bira se samo minimalan podatak (vodostaj ili protok – zavisno o tome koji se problem riješava),*

niz podbačenih vrijednosti ispod praga (treshhold princip) - *iz svake godine biraju se svi oni podaci koji su manji od neke odabrane vrijednosti (tj. praga vodostaja ili protoka)*

U hidrološkim analizama velikih i malih voda česta je potreba **analizirati** i njihova **trajanja**. Razlog za to kod velikih voda su procjene štetnih učinaka poplavnih voda, a kod malih voda razlog leži u procjenama štetnih ekoloških i egzistencijalnih učinaka uslijed pomanjkanja vode.

POVRATNO RAZDOBLJE HIDROLOŠKE VELIČINE

Definicija:

Povratno razdoblje nekog promatranog *(konkretnog)* **iznosa hidrološke** *(ili meteorološke)* **veličine je onaj vremenski period u kojem se taj iznos ili veći od njega pojavljuje prosječno, tj. prema računu vjerojatnosti, jedan puta.**

Tumačenje:

Ako je skup *(tj. vremenska serija)* diskretnih podataka *(tj. mjernih iznosa)* neke hidrološke *(ili meteorološke)* veličine formiran tako da se svaki podatak odnosi na jedno konkretno jedinčno vremensko razdoblje *(najčešće kalendarsku godinu dana)* i ako svakom jediničnom razdoblju *(svakoj godini)* pripada samo jedan reprezentativni podatak tako da je ukupan broj članova skupa podataka jednak ukupnom broju jedinčnih vremenskih razdoblja N iz kojih su podaci prikupljeni *(npr. maksimalni vodostaji iz N godina),* tada vjerojatnost prekoračenja *(tj. kumulativna vjerojatnost)* največeg člana (x_1) promatranog skupa podataka X iznosi

$$F(X \geq x_1) = 1/N$$

(jer ne postoji niti jedan član veći od največeg) pa kažemo da je N povratno razdoblje *(povratni period)* tog člana skupa, tj. $N = 1/\,F(X \geq x_1)$

Da je za najveći član promatranog skupa podataka kumulativna vjerojatnost jednaka: $F(X \geq x_1) = 1/N$ lako se uočava ako promatrani skup podataka sredimo po redosljedu opadanja:

$$\mathbf{x_1 > x_2 > x_3 > \ldots\ldots > x_{N-1} > x_N}$$

N je dakle vremenski period u kojem se promatrani član $\mathbf{x_1}$ iz skupa **X** pojavio jedanput.

Povratno razdoblje obično se označava sa $\mathbf{T_{(x)}}$ *[ili $P_{(x)}$]*. Za prvi *(najveći)* član niza je prema tome: $T\,(x_1) = N = 1/\,F(X \geq x_1)$

Ako dalje promotrimo drugi član po veličini ($\mathbf{x_2}$) u promatranom skupu, taj je član jedan puta dostignut i jedan puta prekoračen u ***N*** jediničnih razdoblja *(godina)*, pa je njegova vjerojatnost prekoračenja jednaka: $F(X \geq x_2) = 2/N = 1/(N/2)$

tj. dva puta je taj član dostignut ili premašen u ***N*** jediničnih razdoblja, odnosno jedan puta u ***N/2*** jediničnih razdoblja.

Dakle za taj je član povratno razdoblje jednako:

$T(x_2) = N/2 = 1/\,F(X \geq x_2)$ jediničnih razdoblja *(godina)*.

Dalje, za treći član po veličini biti će:

$F(X \geq x_3) = 3/N = 1/(N/3)$

$T(x_3) = N/3 = 1/\ F(X \geq x_3)$ jediničnih razdoblja, ... i td.

Općenito vrijedi: $T_{(x)} = 1/\ F(X \geq x)$

Povratno razdoblje je prema tome jednako recipročnoj vrijednosti od vjerojatnosti prekoračenja *(kumulativne vjerojatnosti)*, pod uvjetom da je promatrani skup formiran kako je navedeno *(tj. da se svaki podatak odnosi na jedno konkretno jedinčno vremensko razdoblje, i da svakom jediničnom razdoblju pripada samo jedan reprezentativni podatak tako da je ukupan broj članova skupa podataka jednak ukupnom broju jedinčnih vremenskih razdoblja N)*

Npr. ako je jedinično vremensko razdoblje godina dana i ako za neku vrijednost promatrane hidrološke veličine *(npr. vodostaj Save u Zagrebu od 300 cm)* vjerojatnost prekoračenja iznosi $F(X \geq x) = 0{,}01$ tada povratno razdoblje te vrijednosti *(tj. za tih x=300 cm)* iznosi:

$T_{(x)} = 1/0{,}01 = 100$ godina.

Povratno se razdoblje ustanovljuje računom vjerojatnosti, pa je prema tome to onaj vremenski period u kojem se promatrani iznos hidrološke veličine ili veći od tog promatranog pojavljuje prosječno jedan puta.

To praktično **ne znači da se** u bilo kojem vremenskom periodu, jednako dugačkom kao i povratno razdoblje, **promatrana vrijednost hidrološke veličine mora obavezno i pojaviti.**
Niti to nadalje **ne znači da se** u takovom periodu promatrana vrijednost **ne može pojaviti i dva ili više puta**.
Bit je u tome da se u povratnom razdoblju promatrana vrijednost pojavljuje prosječno jedan puta (npr. *300 cm vodostaja Save u Zagrebu bi se u periodu od recimo 100000 godina prosječno pojavljivalo jedan puta u 100 godina; pri tom bi bilo nekih perioda od 100 godina u kojima se prekoračenje tog vodostaja ne bi dogodilo, ali bi zato bilo i nekih drugih perioda od 100 godina u kojima bi se prekoračenje dogodilo 2 ili više puta)*

Vjerojatnosti prekoračenja koja je po definiciji jednaka recipročnoj vrijednosti povratnog razdoblja odnosi se na svako *(bilo koje)* jedinično razdoblje *(npr. na bilo koju pojedinačnu godinu),* što znači da događaj koji ima ***N*** godišnje povratno razdoblje ima **u svakoj godini** vjerojatnost pojavljivanja ili prekoračenja **$F(X \geq x) = 1/N$**

U hidrološkoj se praksi međutim ne rijetko ukazuje potreba da se proračuna vjerojatnost *(ili rizik)* kojom bi se neki **N** godišnji hidrološki događaj *(koji ima godišnju vjerojatnost pojavljivanja $F(X \geq x) = 1/N$)* mogao pojaviti u nekom vremenskom periodu dužem od jedne godine.

Npr. ako se građevinska jama za izgradnju akumulacijske brane na rijeci štititi izgradnjom zaštitnih nasipa od pojave 100 godišnje velike vode i ako će izgradnja brane trajati 5 godina, postavlja se pitanje kolika je vjerojatnost da se 100 godišnja voda pojavi ili premaši u periodu izgradnje brane?

Zanima nas dakle vjerojatnost prekoračenja u **n** godina, tj. $\mathbf{F(X \geq x)_n = ?}$
Za rješenje ovog zadatka polazi se od izraza:

$$F(X \geq x) + F(X \leq x) = 1 \qquad (a)$$

$$F(X \leq x) = 1 - F(X \geq x) \qquad (b)$$

$F(X \geq x)$ predstavlja godišnju vjerojatnst da će **x** biti dostignuto ili prekoračeno, a $F(X \leq x)$ je godišnja vjerojatnost da **x** ne će biti dostignuto.

Po analogiji vrijedi:

$$F(X \geq x)_n + F(X \leq x)_n = 1 \qquad (c)$$

$$F(X \geq x)_n = 1 - F(X \leq x)_n \qquad (d)$$

$\mathbf{F(X \geq x)_n}$ . . vjerojatnst da će **x** biti jedan puta dostignuto ili prekoračeno u **n** godina;
$\mathbf{F(X \leq x)_n}$ . . vjerojatnost da **x** ne će biti dostignuto niti jednom u **n** godina, *tj.*
*vjerojatnost da će se baš svake godine u periodu od **n** godina desiti podbačaj*

Prema definiciji složene vjerojatnosti: $$F(X \leq x)_n = [F(X \leq x)]^n \qquad (e)$$

uvrštavanjem (b) u (e) biti će: $$F(X \leq x)_n = [1 - F(X \geq x)]^n \qquad (f)$$

uvrštavanjem (f) u (d) : $$F(X \geq x)_n = 1 - [1 - F(X \geq x)]^n \qquad (g)$$

prema definiciji povratnog razdoblja: $$F(X \geq x) = 1/T(x) \qquad (h)$$

uvrštavanjem (h) u (g) biti će: $$F(X \geq x)_n = 1 - [1 - 1/T_{(x)}]^n \qquad (i)$$

U postavljenom primjeru biti će:

$$F(X \geq x)_5 = 1 - [1 - (1/100)]^5 = 0{,}049 = 4{,}90\ \%$$

Dakle, vjerojatnost da 100 godišnja velika voda bude premašena jedan puta u 5 godina iznosi 4,9 % ; za jedan puta u 10 godina vjerojatnost iznosi 9,5 % ; za jedan puta u 50 godina vjerojatnost iznosi 39,5 % ; za jedan puta u 100 godina vjerojatnost iznosi 63,4 %

Proračun povratnog razdoblja i vjerojatnosti prekoračenja ako je niz podataka formiran po principu prekoračenja praga

$$T(x) = \frac{N}{M} \cdot \frac{1}{F(X \geq x)^*} \qquad F(X \geq x) = \frac{M}{N} F(X \geq x)^*$$

T(x) . . . povratno razdoblje
M . . . broj članova niza
N . . . broj godina (*jediničnih vremenskih intervala*)
$F(X\geq x)^*$ . . . vjrojatnost prekoračenja proračunata temeljem ukupnog broja članova niza
$F(X\geq x)$ . . . stvarna vjerojatnost da **x** bude dostignut ili prekoračen u jednoj *(bilo kojoj)* godini *(bilo kojem jediničnom vremenskom intervalu)*

Obrazloženje:
Matematički formalizam proračuna vjerojatnosti prekoračenja ne zna ništa o tome kako je formiran skup podataka pa se proračunata vjerojatnost **$F(X\geq x)^*$** odnosi na ukupan broj podataka, tj. **$F(X\geq x_m)^*$** $= m/M$ S obzirom na to, recipročna vrijednost tako proračunate vjerojatnosti dala bi povratno razdoblje kao da je broj podataka jednak broju godina *(tj. jediničnih vemenskih intervala).* Time bi se dobilo veće povratno razdoblje i manja godišnja vjerojatnost prekoračenja nego u slučaju da je niz podataka formiran po principu maksimuma. Zbog toga je potrebna korekcija izračuna povratnog razdoblja i stvarne godišnje vjerojatnosti prekoračenja na način kako je to dano u gornjim jednadžbama

Primjer proračuna povratnog razdoblja i vjerojatnosti prekoračenja ako je niz podataka formiran po principu prekoračenja praga

Neka ukupan broj članova niza prekoračenja iznosi M=100 a broj godina iz kojih su podaci prikupljeni neka je N=50. Koliko je povratno razdoblje i godišnja vjerojatnost prekoračenja največeg i drugog po veličini člana u zadanom nizu?

Empirijska vjerojatnost proračunata temeljem svih članova niza iznosi:

za naveći član $F(X \geq x_1)^* = 1/100 = 0.01 = 1\ \%$

za drugi član po veličini: $F(X \geq x_1)^* = 2/100 = 0.02 = 2\ \%$

povratno razdoblje navećeg člana: $T(x_1) = (N/M) \cdot 1/ F(X \geq x_1)^* = (50/100) \cdot 1/0{,}01 = 50$ godina

povratno razdoblje drugog člana: $T(x_2) = (N/M) \cdot 1/ F(X \geq x_2)^* = (50/100) \cdot 1/0{,}02 = 25$ godina

stvarna godišnja vjerojatnost prekoračenja najvećeg člana:

$F(X \geq x_1) = (M/N) \cdot F(X \geq x_1)^* = 1/ T(x_1) = (100/50) \cdot 0{,}01 = 1/50 = 0{,}02 = 2\ \%$

stvarna godišnja vjerojatnost prekoračenja drugog člana:

$F(X \geq x_2) = (M/N) \cdot F(X \geq x_2)^* = 1/ T(x_2) = (100/50) \cdot 0{,}02 = 1/25 = 0{,}04 = 4\ \%$

KARAKTERISTIČNI VODOSTAJI

EVV – ekstremno *(katastrofalni)* visoki vodostaj
VVn – **n** godišnji visoki vodostaj
NVV – najviši visoki vodostaj iz nekog razdoblja
$\mathbf{VV_{dmg}}$ – visoki vodostaj nekog datuma
SVV – srednji visoki vodostaj *(mjesečni; godišnji)*
OV – običan vodostaj *(vodostaj 50% trajnosti, medijana)*
SV – srednji vodostaj
FV – najčešći vodostaj *(mod, vodostaj največe učestalosti)*
SNV – srednji niski vodostaj *(mjesečni; godišnji)*
$\mathbf{NV_{dmg}}$ – niski vodostaj nekog datuma
NNV – najniži niski vodostaj iz nekog razdoblja
NVn – **n** godišnji niski vodostaj
ENV – ekstremno *(katastrofalno)* niski vodostaj

KARAKTERISTIČNI PROTOCI

EVV_Q – ekstremno *(katastrofalno)* velika voda
VV_{Qn} – **n** godišnja vielika voda
NVV_Q – najveća velika voda iz nekog razdoblja
VV_{Qdmg} – velika voda nekog datuma
SVV_Q – srednja velika voda *(mjesečna; godišnja)*
OV_Q – obična voda *(protok 50% trajnosti, medijana)*
SV_Q – srednja voda *(srednji protok)*
FV_Q – najčešća voda *(mod, protok največe učestalosti)*
SMV_Q – srednja niska voda *(mjesečna; godišnja)*
MV_{Qdmg} – mala voda nekog datuma
NMV_Q – najmanja mala voda iz nekog razdoblja
MV_{Qn} – **n** godišnja mala voda
EMV_Q – ekstremno *(katastrofalno)* mala voda

SVRHA PRIMJENE RAČUNA VJEROJATNOSTI U HIDROLOGIJI

Osnovni cilj primjene računa vjerojatnosti u hidrologiji je određivanje vjerojatnosti i povratnog razdoblja pojavljivanja neke veličine *(hidrološke ili metorološke)* i to za:

- odabranu veličinu (*iznos*) iz niza izmjerenih podataka;
- za neki veličinu veću od največeg ili manju od najmanjeg izmjerenog podatka. *(kod analiza velikih i malih voda, kod analize jakih kiša, maksimalnih i minimalnih temperatura i td.).*

Za rješavanje prvog zadatka često se primjenjuju empirijske (kompromisne) razdiobe vjerojatnosti (Hazen, Weibull i dr.), a tehnikom ekstrapolacije moguće je pomoću tih razdioba riješavati i drugi spomenuti zadatak;

Za rješavanje drugo-navedenog zadatka češće se primjenjuju teorijske krivulje (*funkcije*) razdiobe vjerojatnosti, ali i prvo navedeni zadatak se sa tim krivuljama uspješno rješava.

Najčešće se u hidrologiji primenjuju **nesimetričene teorijske krivulje razdiobe** vjerojatnosti *(Pearsonove funkcije razdiobe, Gumbelova, Galtonova – log normalana razdioba i dr.)*

Gaussova (normalna) raspodjela se rijeđe koristi za određivanje povratnih razdoblja, ali se koristi za procjene intervala povjerenja.

Gaussova (normalna) **raspodjela**

Gaussova ili normalna raspodjele je *matematički određena, simetrična, dvo-parametarska raspodjela neprekinute slučajne varijable, zvonolika oblika.*

Dobro aproksimira slučajne pogreške.

Raspodjele meteoroloških i hidroloških veličina najčešće nisu simetrične.

Normalna je raspodjela važan čimbenik korelacijskih analiza, a koristi se i za procjene intervala povjerenja, generiranja hidroloških nizova itd.

Međusobne veze varijabli koje imaju normalnu raspodjelu su linearne i vrednuju se na osnovu koeficijenta korelacije.

Kada su raspodjele nesimetrične, podaci se mogu *'normalizirati'* različitim Transformacijama, npr.: $y = x^a$ ili $y = \ln x$, ili $y = \log x$ kojoj odgovara teorijska log-normalna (Galtonova) raspodjela.

Prema Gaussovu zakonu, raspodjela gustoće vjerojatnosti je:

$$p_{(x)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_{sr})^2}{2\sigma^2}}$$

gdje je: $\pi = 3{,}14$; $\boldsymbol{\sigma}$ je standardno odstupanje; e = 2,718 *(baza prirodnog logaritma)*, a $\boldsymbol{x_{sr}}$ je prosjek *(srednja vrijednost)* članova niza.

Supstitucijom u Gaussovu funkciju: $z = \frac{x - x_{sr}}{\sigma}$

dobije se standardna normalna raspodjela: $$p_{(x)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

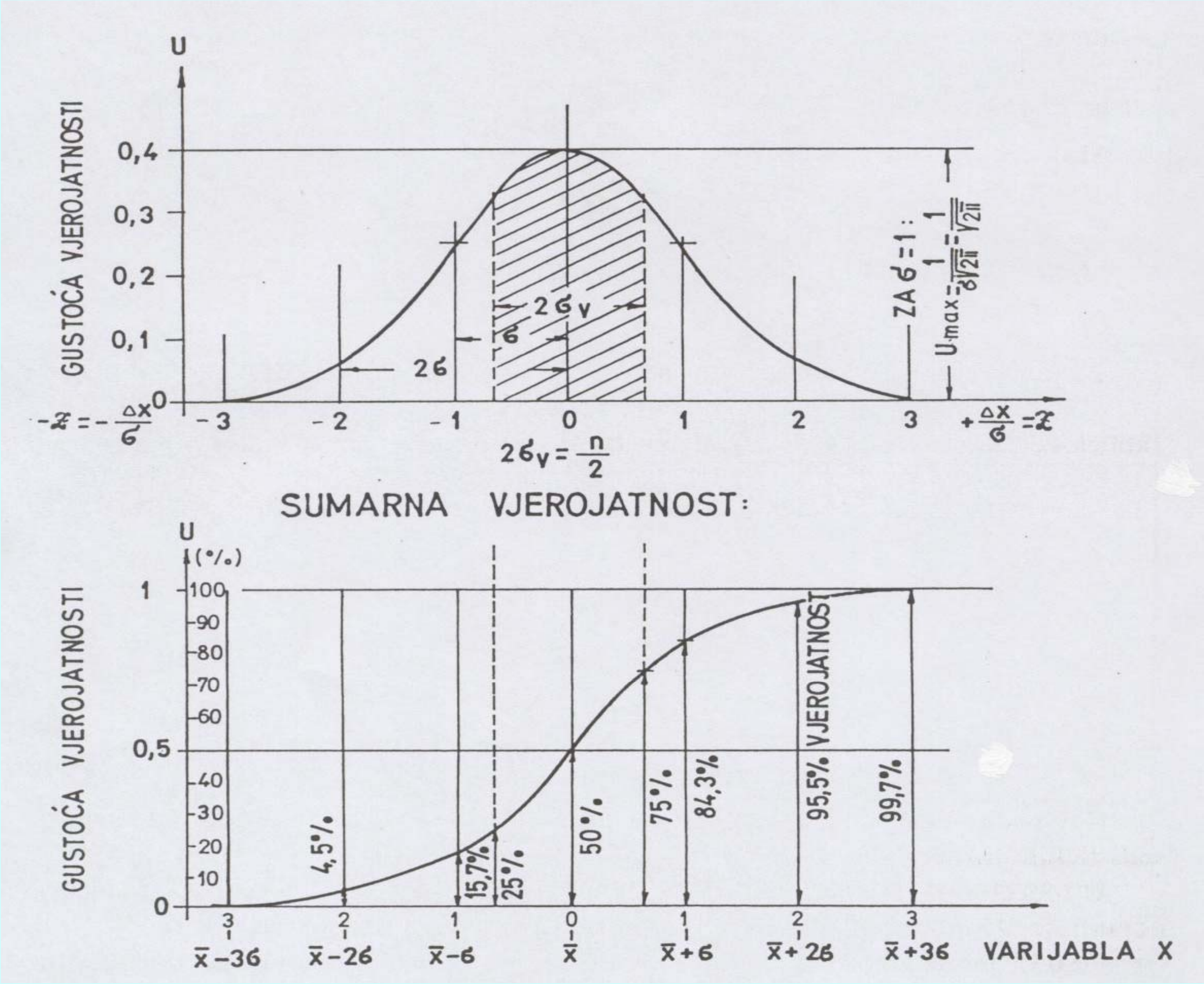
Varijanca standardne normalne raspodjele je $\sigma^2 = 1$, a prosjek *(sredina)* je nula.
Vrijednost $p_{(z)}$ maksimalne ordinate (za $z = 0$ i $\sigma^2 = 1$, tj. $\sigma = 1$) je:

$$p_{(x)} = \frac{1}{\sqrt{2\pi}} = 0{,}399$$

Površina ispod normalne krivulje razdiobe vjerojatnosti jednaka je jedinici.
Ako se ta površina podijeli po sredini na dva jednaka dijela, tada se 50 posto vrijednosti *(tj. članova niza)* nalazi se u intervalu: $\boldsymbol{x_{sr} - 0{,}6745\,\sigma}$ i $\boldsymbol{x_{sr} + 0{,}6745\,\sigma}$
Vjerojatno odstupanje, *unutar kojega se nalazi 50 posto vrijednosti, tj. članova niza,* određeno je izrazom: $\boldsymbol{\sigma_v = \pm\, 0{,}6745\,\sigma}$

Maksimalno odstupanje $\boldsymbol{\sigma_{max}}$ je određeno izrazom: $\boldsymbol{\sigma_{max}} = \pm\, 3\boldsymbol{\sigma}$
U intervalu maksimalnog odstupanja ($\boldsymbol{x_{sr} - 3\sigma}$ i $\boldsymbol{x_{sr} + 3\sigma}$) nalazi se **99,73** posto pojava, tj. gotovo svi članovi razmatranog niza *(skupa)* podataka.

SVOJSTVA NORMALNE KRIVULJE RAZDIOBE VJEROJATNOSTI



Neke karakteristične površine ispod krivulje standardne normalne raspodjele = kumulativne vjerojatnosti

$$F(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz$$

$$F(X \geq x) = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{-\frac{z^2}{2}} dz$$

Ove površine mogu se naći u tablicama statističkih priručnika

Karakteristične površine ispod krivulje standardne normalne raspodjele

z	**F (x_{sr} - zσ < x < x_{sr} + zσ)**
0,6745	0,5000
1	0,6827
1,645	0,9000
1,96	0,9500
2	0,9546
3	0,9973
4	1,0000

INTERVAL I RAZINA POVJERENJA

U hidrologiji se često definira ***vjerojatno relativno odstupanje*** za **interval** ***(raspon)*** **povjerenja** u kojemu se pojavljuje **95 %** podataka.
To odstupanje kod normalne krivulje razdiobe iznsi: $\boldsymbol{\sigma_p = \pm 1{,}96\, \sigma_0}$

$\boldsymbol{\sigma_0}$ je relativno odstupanje:

$$\sigma_0 = \pm\sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i - x_r}{x_r}\right)^2}{n}}$$

gdje je: x_i izmjerena vrijednost *(iznos)* člana niza; x_r je vrijednost izračunata pomoću analitičke krivulje kojom se interpretira stohastička zavisnost niza podataka; n je broj podataka *(članova niza)*
(npr. x_i mogu biti izmjereni protoci Q_i pri raznim vodostajima, a x_r su u tom slučaju računske vrijednosti protoka Q_r prema protočnoj krivulji)

Intervalu povjerenja od **95%** odgovara **razina povjerenja α = ± 5%** *(α = 0,05)*.
Razina povjerenja od **α** = ± 5 % je vjerojatnost da samo 5 % članova razmatranog niza podataka leži izvan intervala povjerenja od 95%, tj. raspona u kojem se nalazi 95% članova niza.

NEKE TEORIJSKE KRIVULJE RASPODJELE VJEROJATNOSTI KOJE SE PRIMJENJUJU U HIDROLOGIJI

BINOMNA RASPODJELA

Diskretna raspodjela dana Bernoulli-*jevom* funkcijom:

$$p(x) = \binom{n}{x} p^x q^{n-x}$$

gdje je:

x - slučajna varijabla koja predstavlja broj pojava *(povoljnih događaja)* u **n** nezavisnih slučajeva *(eksperimenata, godina)*

p - vjerojatnost da se desi neki događaj **x** pri jednom pokušaju *(godini)*

q - vjerojatnost nepojavljivanja događaja **x** pri jednom pokušaju: q = 1 – p

n - ukupan broj nezavisnih slučajeva *(eksperimenata, godina)*

$\binom{n}{x}$ - broj kombinacija u **n** slučajeva od kojih se **x** promatra u isto vrijeme, tj. broj načina da se **x** događaja može pojaviti u **n** slučajeva *(godina)*

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{n(n-1)...(n-x+1)}{x!}$$

p(x) - vjerojatnost da će se u nizu od **n** slučajeva (*godina*) neki događaj desiti **x** puta

srednja vrijednost: x = p·n ; standardna devijacija: $\sigma = \sqrt{p \cdot q \cdot n}$

koeficijent asimetrije: $C_s = \dfrac{q-p}{\sqrt{p \cdot q \cdot n}}$

Primjer određivanja vjerojatnosti pojave prekoračenja vodostaja

U razdoblju od N = 40 godina mjerenja na vodomernoj postaji u Zagrebu zabilježeno je 28 slučajeva *(u 28 god.)* prekoračenja visokog vodostaja od H = 300 cm.

1.) Kolika je empirijska vjerojatnost **p** da se prekoračenje dogodi, a kolika vjerojatnost **q** da se prekoračenje ne dogodi u jednoj *(bilo kojoj)* godini? :

$$\mathbf{p} = F(X \geq x) = 28/40 = 0{,}7 \quad ; \quad \mathbf{q} = F(X \leq x) = 1 - p = 1 - 0{,}7 = 0{,}3$$

2.) Koliko je povratno razdoblje prekoračenja takovog događaja :

$$\mathbf{T_R = 1/p = 1/0{,}7 = 1{,}43 \approx 1{,}5\ god.,}$$

tj. prekoračenje se prosječno događa 1 puta u 1,5 godina

3.) Kolika je vjerojatnost da će se u razdoblju od n = 5 godina prekoračenje dogoditi najmanje jedan put? :

$$\mathbf{p_n = F(X \geq x)_n = 1 - [1 - F(X \geq x)]^n = 1 - [1 - 0{,}7\,]^5 = 0{,}99757 \approx 1}$$

U razdoblju od n = 5 god. prekoračenje vodostaja od H = 300 cm će se gotovo sigurno dogoditi najmanje 1 put.

Primjer određivanja vjerojatnosti pojave prekoračenja vodostaja - *nastavak*

Kolika je vjerojatnost da se u razdoblju od 5 god. prekoračenje: ne dogodi, dogodi upravo 1 put, dogodi 2 put, 3 put, 4 puta, 5 puta?:

Prekoračenje će se dogoditi	varijabla	izračun prema binomnoj razdiobi
nula puta *(ne će se dogditi)*	x = 0	$p_0 = \binom{n}{x} p^x q^{n-x} = \binom{5}{0} \cdot 0{,}7^0 \cdot 0{,}3^5 = 0{,}0024 = 0{,}24\%$
upravo *(samo)* jedan put	x = 1	$p_1 = \binom{n}{x} p^x q^{n-x} = \binom{5}{1} \cdot 0{,}7^1 \cdot 0{,}3^4 = 0{,}02835 = 2{,}84\%$
upravo dva put	x = 2	$p_2 = \binom{n}{x} p^x q^{n-x} = \binom{5}{2} \cdot 0{,}7^2 \cdot 0{,}3^3 = 0{,}1323 = 13{,}23\%$
upravo tri put	x = 3	$p_3 = \binom{n}{x} p^x q^{n-x} = \binom{5}{3} \cdot 0{,}7^3 \cdot 0{,}3^2 = 0{,}3087 = 30{,}87\%$
upravo četiri puta	x = 4	$p_4 = \binom{n}{x} p^x q^{n-x} = \binom{5}{4} \cdot 0{,}7^4 \cdot 0{,}3^1 = 0{,}3602 = 36{,}02\%$
upravo pet puta	x = 5	$p_5 = \binom{n}{x} p^x q^{n-x} = \binom{5}{5} \cdot 0{,}7^5 \cdot 0{,}3^0 = 0{,}1681 = 16{,}81\%$

POISSONOVA RASPODJELA

Poissonova raspodjela (granični slučaj binomne raspodjele): $p(x) = \dfrac{m^x \cdot e^{-m}}{x!}$

Statistički parametri:

sredina: $x_{sr} = m$; standardno odstupanje: $\sigma = m^{0,5}$

koeficijent asimetrije: $C_s = 1 / m^{0,5}$ mjera spljoštenosti: $3 + C_s$

Za razne vrijednosti od **x** biti će **p(x)** jednako:

$x = 0$: $p(0) = e^{-m}$

$x = 1$: $p(1) = m \cdot e^{-m}$

$x = 2$: $p(2) = (½) \cdot m^2 \cdot e^{-m}$

$x = 3$: $p(3) = (1/6) \cdot m^3 \cdot e^{-m}$ i td.

Poissonova raspodjela naziva se još ‘zakonom malih brojeva’. Koristi se za slučajeve kada je ***N*** vrlo velik a vjerojatnost ***p = m/N*** vrlo mala, dakle za analizu hidroloških veličina s velikim povratnim razdobljima i malim vjerojatnostima pojavavljivanja.

Primjer: kolika je vjerojatnost da se 1000 godišnja voda pojavi u sljedečih 100 godina? :

$m = n / T = 100 / 1000 = 0,1$

- da se uopće ne dogodi : $x = 0$: $p(0) = e^{-m} = e^{-0,1} = 0,9048 = 90,48$ %
- da se dogodi 1 put : $x = 1$: $p(1) = m \cdot e^{-m} = 0,1 \cdot e^{-0,1} = 0,09048 = 9,048$ %
- da se dogodi 2 put : $x = 2$: $p(2) = (½) \cdot m^2 \cdot e^{-m} = (½) \cdot 0,1^{2} \cdot e^{-0,1} = 0,0045 = 0,45$ %
- da se dogodi 3 put : $x = 3$: $p(3) = (1/6) \cdot m^3 \cdot e^{-m} = (1/6) \cdot m^3 \cdot e^{-0,1} = 0,00015 = 0,015$%

Log-normalna (Galtonova) **raspodjela**

Galton je dokazao da se mnoge raspodjele vjerojatnosti mogu u logaritamskoj transformaciji prilagoduti Gaussovom zakonu:

$$p_{(x)} = \frac{1}{\sigma_y \sqrt{2\pi}} \cdot e^{-\frac{(y-y_{sr})^2}{2\sigma_y^2}}$$

pri čemu je:

y = ln x (**x** je varijabla)

$\mathbf{y_{sr}}$ .... sredina od **y**

$\boldsymbol{\sigma_y}$ .... standardna devijacija od **y**

Između parametara originalnog niza podataka x_{sr} i σ_x , i parametara izvedenog niza y_{sr} i σ_y

postoji odnos: $y_{sr} = \ln \frac{x_{sr}^2}{\sqrt{\sigma_x^2 + x_{sr}^2}}$ i $\sigma_y = \sqrt{\ln(1 + \frac{\sigma_x}{x_{sr}^2})}$

Statistički parametri Galtonove raspodjele su:

koeficijent varijacije: $C_v = (e^{\sigma_y^2} - 1)^{1/2}$

koeficijent asimetrije: $C_s = 3C_v + C_v^3$

medijana: $x_{me} = e^{y_{sr}}$

PEARSONOVE FUNKCIJE RASPODJELE

Opća (osnovna) jednadžba gustoće vjerojatnosti: $$p(x) = e^{\int_{-\infty}^{x} (m-x)(a+bx+cx^2)\,dx}$$

Gdje su ***a, b, c*** i ***m*** konstante. Kriteriji za određivanje tipova funkcija su β_1, β_2 i k :

$$\beta_1 = \frac{\eta_3^2}{\eta_2^2} \qquad \beta_2 = \frac{\eta_4^2}{\eta_2^2} \qquad k = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)\cdot(2\beta_2-3\beta_1-6)}$$

η_2 , η_3 i η_4 su drugi, treći i četvrti centralni momenti.
Osnovna jednadžba ima 4 parametra (***a, b, c*** i ***m***), pa se prema njihovim vrijednostima dobivaju različite krivulje. Sve su te krivulje unimodalne, tj. imaju jedan vrh (mod) u točki $x = m$. Zbog toga ove krivulje imaju veliku sposobnost prilagođavanja. Postoji 12 tipova tih krivlja, a u hidrologiji se najčešće koristi III tip Pearsonove raspodjele kod koje je:
$k = \infty$, odnosno $2\beta_2 = 3\beta_1 + 6$, a gustoća vjerojatnosti je:

$$p(x) = p_0\left(1+\frac{x}{m}\right)^c \cdot e^{-c\frac{x}{2}}$$

gdje je:

$$c = \frac{4}{\beta_1} - 1 \qquad m = \frac{c}{2}\cdot\frac{\eta_3}{\eta_2} \qquad p_0 = \frac{N}{m}\cdot\frac{c^{c+1}}{e^c\cdot\Gamma(c+1)} \qquad \Gamma(c+1) = \int_0^{\infty} x^c e^{-x}\,dx = c!$$

Ova se Pearsonova raspodjela često koristi u modifikaciji Fostera i Ribkina.
Više vidi: *Prof. Dionis Srebrenović: Primjena matematsko statističkih metoda u hidrologiji, Sveučilište u Zagrebu, 1970.*

JEDNOPARAMETARSKA GAMA RASPODJELA

Gustoća vjerojatnosti za ovu raspodjelu je: $$p(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \cdot e^{-x}$$ $\boxed{\text{za } x \geq 0}$

$p(x) = 0$ za $x < 0$

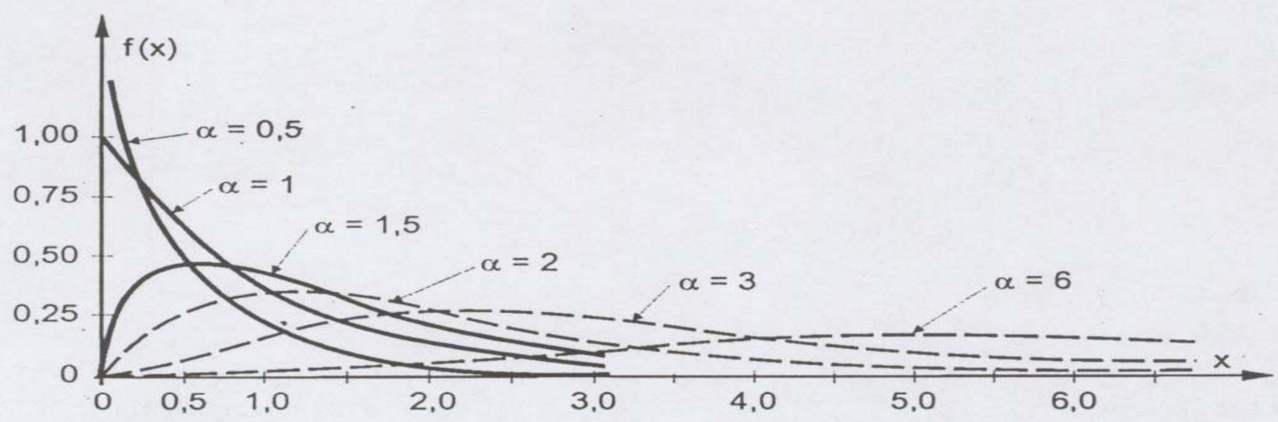
gdje je: $$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \cdot e^{-x} dx$$

Integral $\Gamma(\alpha)$ konvergira ako je $x > 0$ i zove se gama funkcija od α.

Vrijedi: $\Gamma(1) = 1$; $\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1)$; $\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha)$

$\Gamma(\alpha) = (\alpha - 1)!$; $\Gamma(\alpha+1) = \alpha!$

$x_{sr} = \alpha$; $\sigma = \alpha^{1/2}$; $C_v = \alpha^{-1/2}$; $C_s = 2 \cdot \alpha^{-1/2}$



Funkcija gustine raspodjele vjerovatnoće jednoparametarske Gama raspodjele za različite vrijednosti α

DVOPARAMETARSKA GAMA RASPODJELA

Gustoća vjerojatnosti za ovu raspodjelu je: $$p(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \cdot e^{-\frac{x}{\beta}}$$

$0 \leq x \leq \infty \quad \alpha > 0 \quad \beta > 0$

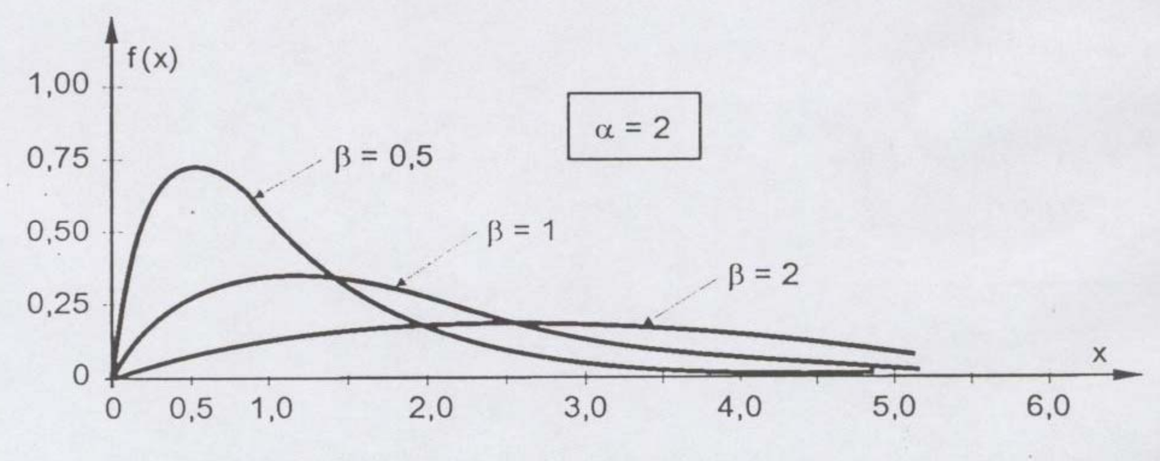
$p(x) = 0$ za $x < 0$

gdje je: $$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \cdot e^{-x} dx$$

Kumulativna funkcija vjerojatnosti je: $$F(x) = \int_0^{x} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\frac{x}{\beta}} dx$$

Za dato α i β gornji integral F(x) riješava se numeričkom integracijom

Karakteristične vrijednosti raspodjele: $x_{sr} = \alpha \cdot \beta$; $\sigma = \beta \cdot \alpha^{1/2}$; $C_v = \alpha^{-1/2}$; $C_s = 2 \cdot \alpha^{-1/2}$



Funkcija gustine raspodjele vjerovatnoće dvoparametarske Gama raspodjele za karakteristične vrijednosti β i α = 2