Kodiranje medijskih signala

Prof.dr.sc. Davor Petrinović,

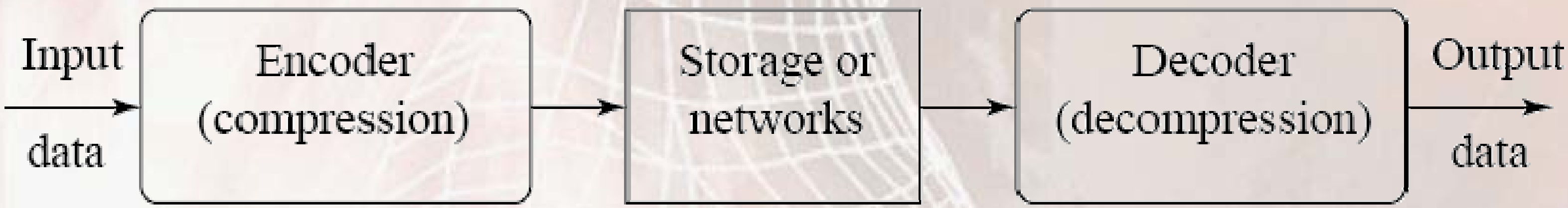
20.veljače 2008.

Kodiranje medijskih signala

- Manipulacija sa digitaliziranim medijskim signalima u formatu nominalne točnost nije pogodna.
- Takav format je previše rastrošan:
  - traži veliki podatkovni tok, dok je
  - stvarna količina informacije sadržana u signalu nekoliko puta manja, a ponekad čak i nekoliko redova veličine puta manja!
- Potrebno je osmisliti postupak kojim je moguće **odvojiti bitno od nebitnog**.

Kodiranje medijskih signala

- Općenito, ... kodiranje (engl. ***coding***) se definira kao predstavljanje informacije nizom bita.
- Kompresija ili sažimanje (engl. ***data compression***) je:
  - kodiranje kojem je cilj smanjiti broj bita kojim se prikazuje određena informacija.



Kodiranje medijskih signala

- Često se termin kodiranja koristi kao sinonim za postupke kompresije, što će biti slučaj i u našim materijalima.
- Kompresija se još naziva postupkom uklanjanja zalihosti, engl. ***redundancy removal***.
- Zadatak kompresije je:
  - format nominalne točnosti transformirati u "ekvivalentu" digitalnu reprezentaciju u kojoj je smanjena ili uklonjena zalihost informacije.

Kodiranje medijskih signala

- Kodiranje u širem smislu se može podijeliti u tri podvrste:
  – kodiranje izvora (engl. ***source coding***),
  – entropijsko kodiranje (engl. ***entropy coding***) i
  – kodiranje kanala (engl. ***channel coding***).
- Prva dva se koriste u kontekstu kompresije, dok se, ...
  – kod kodiranja kanala zapravo **povećava** broj bita poruke u svrhu zaštite informacije prilikom prijenosa.

Kodiranje medijskih signala

- U okviru predmeta MT, baviti ćemo se prvenstveno sa kodiranjem izvora i entropijskim kodiranjem.
- Kodiranje dijelimo i na:
  - kodiranje bez gubitaka (engl. ***Lossless coding***),
  - kodiranje s gubitcima (engl. ***Lossy coding***).
- Savršena rekonstrukcija signala u formatu nominalne točnosti moguća je samo u slučaju kodiranja bez gubitaka!

Kodiranje medijskih signala

- Koja je razlika između kodiranja izvora i entropijskog kodiranja?
  - kodiranje izvora je temeljeno na svojstvima signala na koji se primjenjuje i iskorištava zalihosti koje su vezane uz strukturu konkretnog signala, dok ...
  - entropijsko kodiranje “napada” digitalnu poruku bez poznavanja njenog pravog izvora i ostvaruje kompresiju na račun statističkih svojstava tog digitalnog niza uzoraka.

Informacija i entropija

prof.dr.sc. Davor Petrinović

Priprema materijala:

Ivan Dokmanić, dipl.ing.

Informacija

- Informaciju (količinu informacije) možemo definirati kao **razinu iznenađenja zbog nekog događaja**, odnosno kao nevjerojatnost tog događaja.
- Primjer – pogađanje ženskog imena:
  – Ako nam netko kaže kako je prvo slovo imena *M* (često, vjerojatno prvo slovo), to nam ne znači mnogo, ali ako nam kaže da ime počinje sa *Ž*, količina informacije je mnogo veća i lakše ćemo pogoditi o kojem se imenu radi.

Mjera informacije

- Pretpostavimo bezmemorijski izvor informacija koji može generirati diskretne simbole iz skupa: $S = \{s_1, s_2, s_3, \ldots, s_n\}$
- Ako je vjerojatnost generiranja *i*-tog simbola p_i, količina informacije zbog opažanja tog *i*-tog simbola je:

$$\log_2 \frac{1}{p_i}$$

- Uzimanje logaritma po bazi 2 daje rezultat u bitovima – (minimalan) broj bitova potreban za reprezentaciju simbola s_i

Entropija

- Entropija nekog izvora informacije je **mjera količine informacije** koju taj izvor generira.
- Ona odgovara prosječnoj (očekivanoj) količini informacije po generiranom simbolu:

$$H(S)=\sum_{i=1}^{n} p_i \log\frac{1}{p_i}$$

$$H(S)=-\sum_{i=1}^{n} p_i \log p_i$$

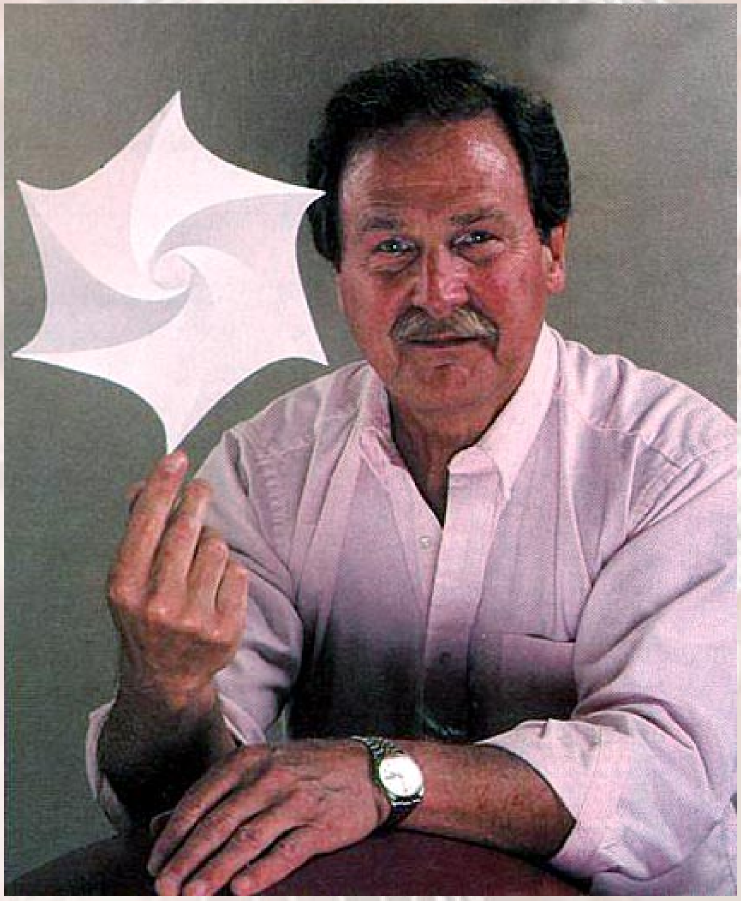
Entropija

- Prethodno definirana entropija odgovara najmanjoj mogućoj prosječnoj dužini kodirane poruke u bitovima.
- Pri tome pretpostavljamo da je izvor bezmemorijski, tj. da vjerojatnost pojave trenutnog simbola niti na koji način nije vezana sa pojavom bilo kojeg drugog simbola prije razmatranog.

Kodovi promjenjive duljine

- Ako su vjerojatnosti pojave pojedinih simbola različite, dobra je ideja onima vjerojatnijima (koji se češće pojavljuju) pridijeliti kraće kodne riječi i tako smanjiti očekivanu dužinu kodirane poruke.
- Kako su kodne riječi različitih dužina, **niti jedna ne smije biti početak neke druge**, radi ostvarenja mogućnosti jedinstvenog dekodiranja.
- Kodovi koji imaju to svojstvo zovu se **prefiksni kodovi**.

Huffmanovi kodovi



- David A. Huffman (1925-1999)

Huffmanovi kodovi

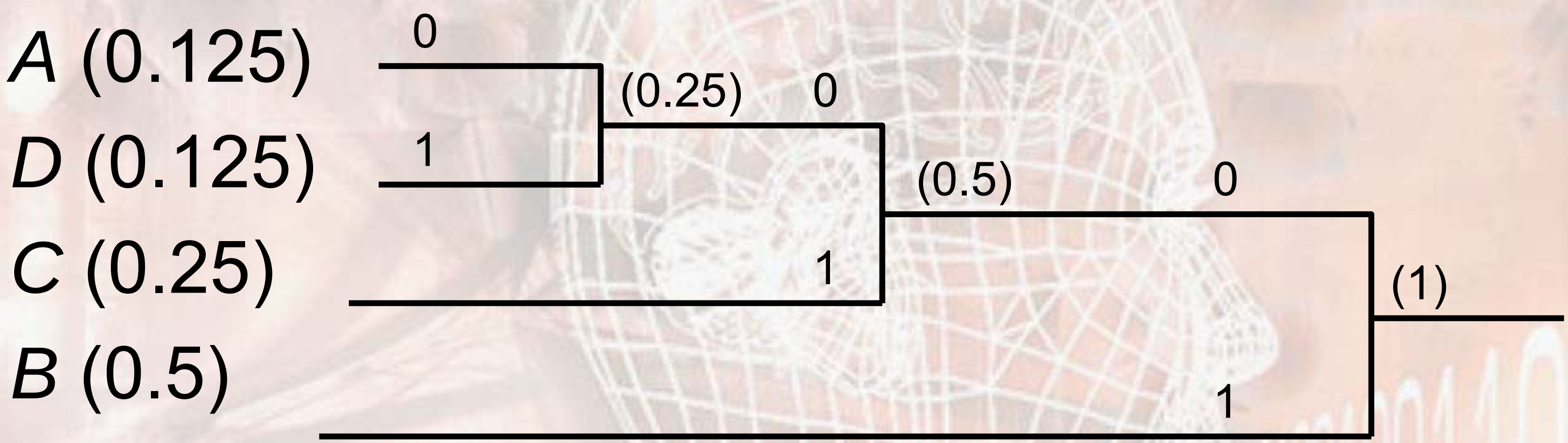
- Zadovoljavaju uvjet prefiksa!
- Algoritam za formiranje Huffmanovog koda:
  1. Sortiraj simbole prema vjerojatnosti,
  2. Odaberi 2 simbola najmanje vjerojatnosti, napravi Huffmanovo pod-stablo čija su djeca ti simboli,
  3. Roditelju dodijeli vjerojatnost koja je zbroj vjerojatnosti djece i ubaci ga u listu simbola tako da je ona i dalje sortirana, a djecu izbaci iz liste
  4. Ponavljaj od 2 dok ne iskoristiš sve simbole,
  5. Dodijeli kodne riječi svakom listu ovog stabla (one odgovaraju putovima od korijena stabla ka listu).

Huffman – primjer

- Uzmimo izvor čiji je alfabet $S = \{A, B, C, D\}$
- Konstruirajmo Huffmanov kod za taj izvor, ako su vjerojatnosti pojedinih simbola kao u tablici

A	*B*	*C*	*D*
0.125	0.5	0.25	0.125

Huffman – primjer



Kodne riječi čitamo od korijena stabla prema listovima

A	*B*	*C*	*D*
000	1	01	001

Huffman – primjer

- Prosječna dužina kodne riječi:

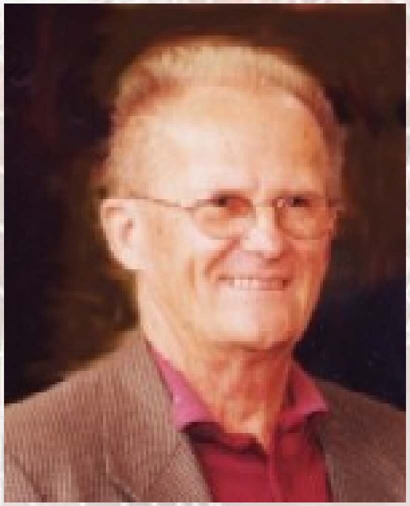
  0.125*3 + 0.125*3 + 0.25*2 + 0.5*1 = 1.75 bita
- Ta dužina **točno odgovara** entropiji izvora!
- Kodirana poruka **BADBBCBC** je: **10000011101101** i dužine je 14 bita. Da smo koristili jednaku dužinu kodne riječi za sve simbole (2 bita), trebali bismo ukupno 16 bita.
- Da vjerojatnost svakog simbola nije bila cjelobrojna potencija od 0.5 (odnosno da količina informacije po simbolu nije bila cjelobrojna), prosječna dužina kodne riječi bila bi **do najviše za 1 bit veća od entropije**!

Huffman - problem

- Za prirodne procese, količina informacije po simbolu uistinu ne mora biti cjelobrojna, a ova metoda svakom simbolu nužno dodjeljuje cijeli broj bita.
- Ekstreman slučaj je kada je vjerojatnost nekog simbola vrlo velika (>0.5), npr. 0.9
- Količina informacije tog simbola je 0.152 bita, ali će njegova kodna riječ biti dužine 1 bit i prosječna dužina koda značajno veća od entropije.

Aritmetičko kodiranje

- Kako ostvariti frakcionalnu dužinu koda ?
- Ideja: cijelu poruku kodirati odjednom – tj. cijelu poruku kodirati jednim “dugačkim” brojem.



- Jorma J. Rissanen
- smislio postupak dok je radio za IBM ...

Aritmetičko kodiranje

- Neka je $S = \{A, B, C, D, E, F, \$\}$, s vjerojatnostima kao u tablici
- Svakom simbolu dodijelimo podinterval od [0,1), čija širina odgovara vjerojatnosti simbola:
  - česti simboli – širok interval,
  - rijetki simboli – uzak interval.
- Redoslijed intervala je proizvoljan

A	B	C	D	E	F	$
0.2	**0.1**	**0.2**	**0.05**	**0.3**	**0.05**	**0.1**
[0, 0.2)	**[0.2,0.3)**	**[0.3,0.5)**	**[0.5,0.55)**	**[0.55,0.85)**	**[0.85,0.9)**	**[0.9,1)**

Aritmetičko kodiranje

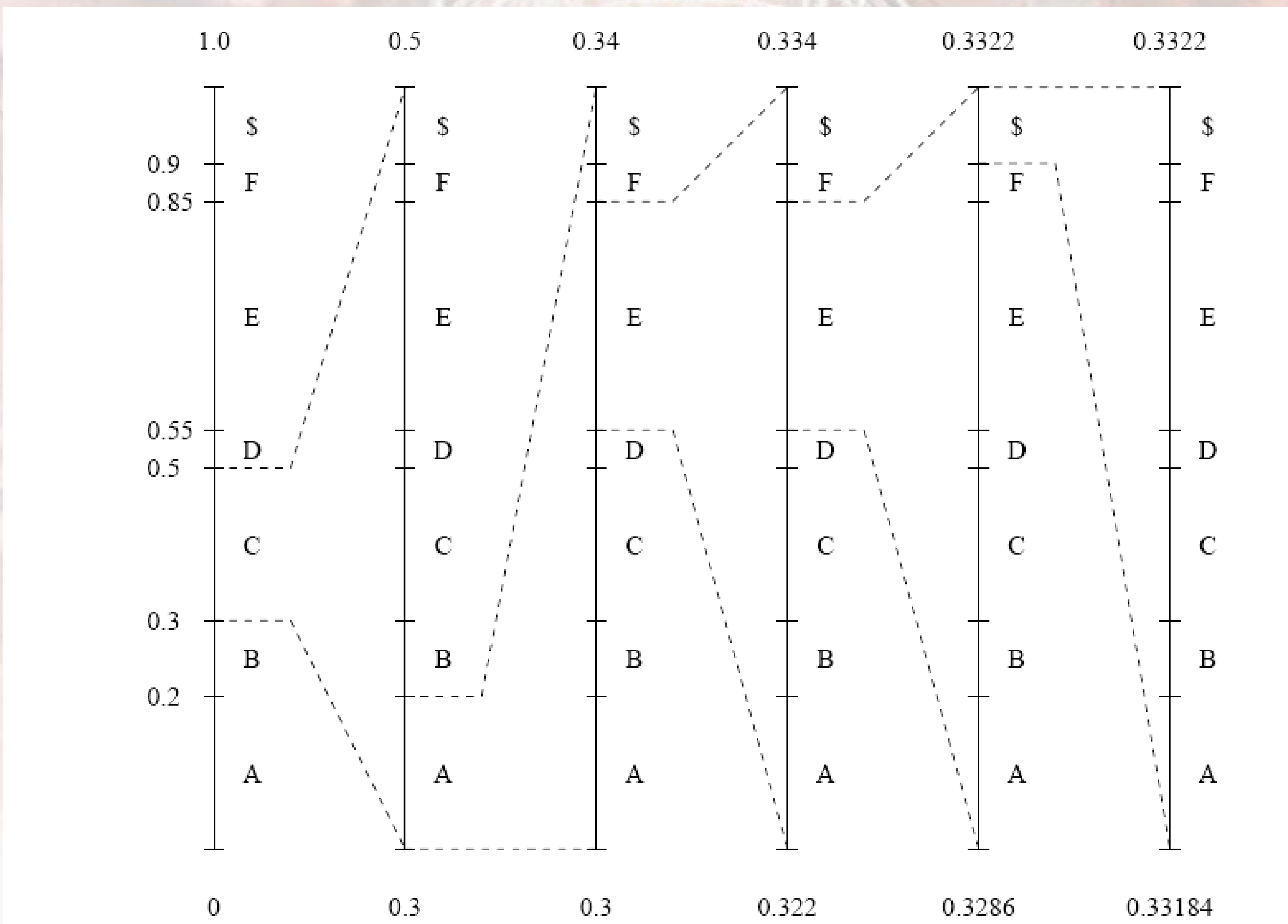
- Kada kodiramo prvi simbol, koncentriramo se samo na podinterval koji odgovara tom simbolu, i njega podijelimo jednako kao što smo podijelili [0,1), tako da narednim simbolima odgovaraju podintervali unutar njega, a širine podintervala su opet proporcionalne vjerojatnostima simbola.
- Kad kodiramo sljedeći simbol, istu stvar napravimo sa podintervalom koji odgovara tom simbolu (a nalazi se unutar intervala koji odgovara prethodnom simbolu).
- Postupak rekurzivno ponavljamo do kraja poruke, a kodirana poruka je bilo koji broj iz završnog intervala

Aritmetičko kodiranje - primjer

- Neka je poruka koju kodiramo CAEE$
- Postupak je ilustriran sljedećom tablicom u kojoj je za svaki simbol poruke naveden interval koji promatramo:

Simbol	Gornja	Donja	Širina
	0	1	1
C	0.3	0.5	0.2
A	0.3	0.34	0.04
E	0.322	0.334	0.012
E	0.3286	0.3322	0.0036
$	0.33184	0.3322	0.00036

Aritmetičko kodiranje – primjer



Aritmetičko kodiranje – primjer

- Na kraju je potrebno odabrati najkraću moguću kodnu riječ, odnosno broj iz završnog intervala sa najmanje znamenaka.
- Kada bi podatke predstavljali dekadskim, a ne binarnim znamenkama, taj bi broj očigledno bio 0.332, no u bazi 2 taj broj ima beskrajno mnogo znamenaka.
- Kako u binarnom brojevnom sustavu imamo tek 2 mogućnosti za svaku znamenku, jednostavno je konstruirati algoritam koji će generirati binarni broj s najmanje znamenaka u nekom intervalu.

Aritmetičko kodiranje

1. Postavi *b* na nulu, a brojač *i* na 1
2. Dokle god je *b* manji od donje granice intervala
   1. Postavi *i*-tu binarnu frakcijsku znamenku od *b* na 1
   2. Ako je *b* veći od gornje granice intervala, vrati *i*-tu znamenku na nulu,
   3. Uvećaj brojač *i* za 1
3. Uvjet završetka petlje biti će objašnjen kasnije!

- Završni frakcionalni broj *b* odgovara traženom broju, i predstavlja kod originalne poruke.
- U našem primjeru se radi o broju **0.33203125**
  - odnosno u binarnom prikazu **.01010101**

Aritmetičko kodiranje – zašto radi

- Širina završnog intervala odgovara upravo produktu vjerojatnosti svih odaslanih simbola.
- Ako simbole smatramo nezavisnima, tada je taj produkt baš vjerojatnost cijele poruke, p_m.
- Općenito, ako želimo prikazati brojeve iz jako uskih intervala unutar [0, 1), morat ćemo trošiti mnogo znamenaka, jer uski intervali odgovaraju malim vjerojatnostima p_m.

Aritmetičko kodiranje – zašto radi

- U bilo kojem intervalu širine p_m unutar [0,1) postoji realni broj koji se može prikazati binarnom frakcijom čiji je broj binarnih znamenaka najmanji cijeli broj veći od $\log_2(1/p_m)$.
- Veličina $\log_2(1/p_m)$ je upravo količina informacije u cijeloj poruci, što znači da smo poruku kodirali najmanjim mogućim brojem bita, tj. sa najviše do jednog bita viška za cijelu poruku!
- Kod Huffmana svaki simbol kodiramo sa do bitom viška!

Aritmetičko kodiranje – zašto radi

- Za naš primjer, vjerojatnost poruke CAEE$ je:
  – $p_m = p_C\, p_A\, p_E\, p_E\, p_\$ = 0.2 \cdot 0.2 \cdot 0.3 \cdot 0.3 \cdot 0.1 =$
  – $p_m = 0.00036$
- Minimalni broj bita za cijelu poruku bio bi jednak entropiji:
  – $-\log_2(p_m) = 11.44$ bit
- U realnom sustavu koristimo 12 bita i šaljemo poruku:
  – **.010101010000**
- Dakle, broj prolaza petlje za traženje frakcije je 12

Aritmetičko kodiranje

- Proces dekodiranja je jednostavan i svodi se na čitanje slike sa podjelom pod-intervala unazad:
  - Pogledamo unutar kojeg intervala se nalazi trenutni broj (kodna riječ);
  - trenutni dekodirani simbol odgovara tom intervalu.
  - Sljedeći je korak oduzimanje donje granice identificiranog intervala od trenutne vrijednosti broja i skaliranje na jedinični interval dijeljenjem sa širinom intervala (tj. vjerojatnosti) dekodiranog simbola.
  - Tako modificirani broj vraćamo u prvi korak algoritma i ponavljamo postupak.

Aritmetičko kodiranje

- Ilustracija dekodiranja poruke, opisane 12-bitnom frakcijom *b*= **.010101010000**

Trenutna vrijednost	Izlazni simbol	Donja granica	Gornja granica	Širina interval a
0.33203125	C	0.3	0.5	0.2
0.16015625	A	0	0.2	0.2
0.80078125	E	0.55	0.85	0.3
0.8359375	E	0.55	0.85	0.3
0.953125	$	0.9	1	0.1

Aritmetičko kodiranje

- Potencijalni problem nastaje zbog činjenice da se proces dekodiranja može nastaviti do beskonačnosti.
- To se rješava uvođenjem simbola koji simbolizira kraj poruke ($ u primjeru).
- Uvođenjem tog simbola narušava se optimalnost, ali bez obzira na to ovaj je postupak u većini slučajeva superioran Huffmanovom kodiranju.
- Jedan od razloga slabije primjene ovog algoritma je zbog činjenice da je zaštićen većim brojem patenata!

Aritmetičko kodiranje – problem

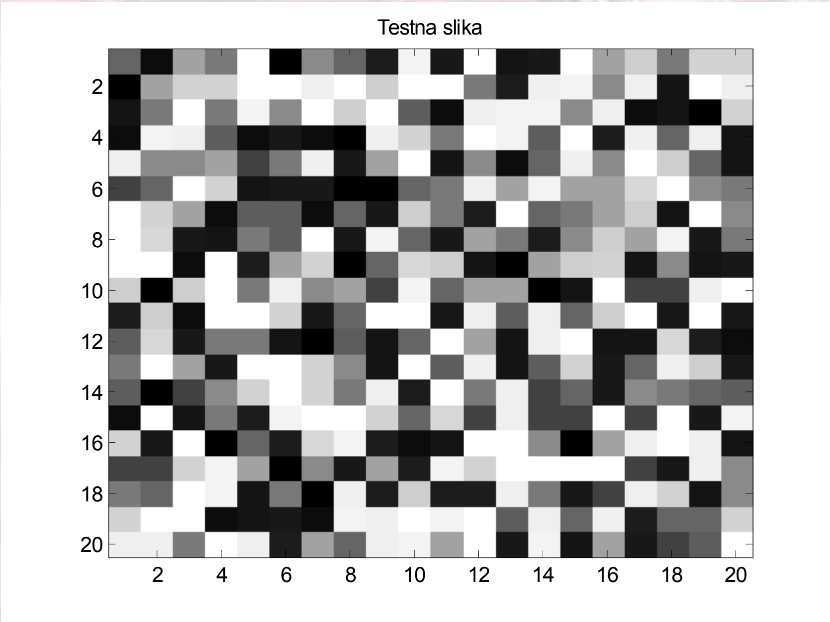
- Da je naša idealna kodirana poruka pogreškom izmijenjena i to tako da smo samo zadnji bit poruke podigli u 1
  - **.010101010001** (0.332275390625)
- ... to bi imalo kobnu posljedicu na dekodiranje, jer je ona izvan završnog intervala kodiranja [0.33184,0.3322) i dekodira se kao:
  - CAEFAEBEA$ te odgovara 26-bitnoj frakciji:
  - **.01010101000100000000000000**

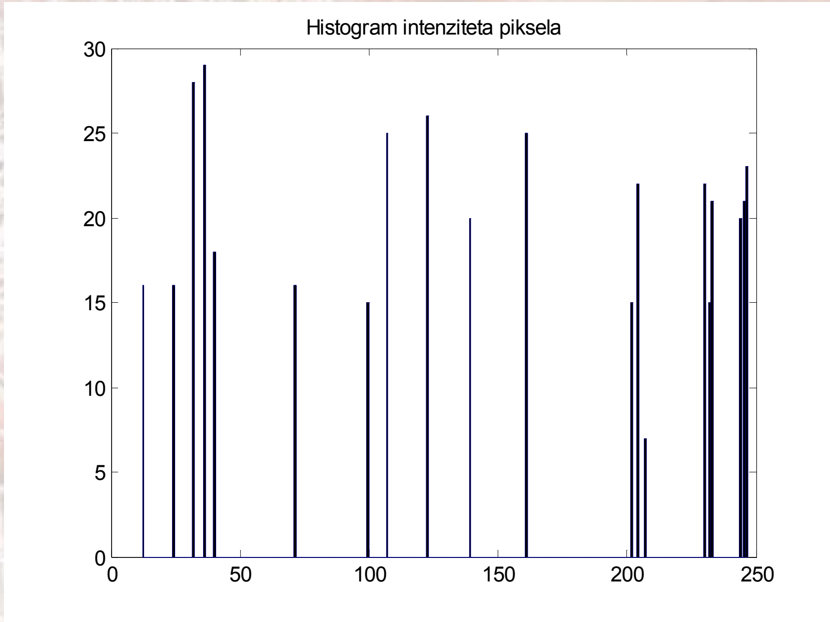
Aritmetičko kodiranje – problem

- Pogrešno dekodirana poruka:

Trenutna vrijednost	Izlazni simbol	Donja granica	Gornja granica	Širina intervala
0.332275390625	C	0.3	0.5	0.2
0.161376953125	A	0	0.2	0.2
0.806884765625	E	0.55	0.85	0.3
0.85628255208333	F	0.85	0.9	0.05
0.12565104166668	A	0	0.2	0.2
0.62825520833339	E	0.55	0.85	0.3
0.26085069444465	B	0.2	0.3	0.1
0.60850694444649	E	0.55	0.85	0.3
0.19502314815497	A	0	0.2	0.2
0.97511574077485	$	0.9	1	0.1

Primjer primjene entropijskog kodera na sliku





Entropije slike je 4.2688 bita po pixelu, dakle minimalan broj bita za cijelu sliku je 20*20*4.2688 = 1707.5, za razliku od 20*20*8 = 3200 bez kompresije.

Zaključak

- Huffmanov koder i aritmetički koder su oba entropijski koderi:
  - sažimanje ostvaruju primjenom koda varijabilne dužine u skladu sa statistikom signala;
  - time se skraćuje prosječna dužina koda i izlazni podatkovni tok kodirane poruke približava teorijskom minimumu definiranom entropijom;
  - dodatna prednost aritmetičkog koda je uslijed frakcionalne dužine koda pojedinog simbola.

Što smo naučili

- definicija kodiranja i kompresije
- tipovi kodiranja i podjele
- informacija
- mjera informacije, entropija
- entropijski koderi
- kodovi promjenjive duljine
- Huffmanovi kodovi
- aritmetički koder

Entropija i kvaliteta

prof.dr.sc. Davor Petrinović

Kodiranje izvora radi kompresije

- Problem:
  - Imamo amplitudno kontinuirani vremenski diskretni proces, kojeg želimo kodirati u svrhu kompresije;
  - amplitudno kontinuirani proces pretvaramo u diskretni digitalizacijom (kvantizacijom);
  - želimo primijeniti entropijski koder na izlazu kvantizatora, radi kompresije.
  - Koji je odnos kvalitete (pogreške kvantizacije) i entropije izlazne kodirane poruke ?

Kodiranje izvora radi kompresije

- Kodiranje **s ograničenom entropijom**
  – Kanal ima konačni raspoloživi podatkovni tok kroz koji se šalje kodirana informacija.
  – Kako projektirati kvantizator da nakon entropijskog kodiranja poruka “stane” u kanal, a da se pri tome ostvari čim veća kvaliteta?
- Kodiranje **s ograničenom distorzijom**
  – Rekonstruirani signal mora imati pogrešku kvantizacije manju od propisane.
  – Kako projektirati kvantizator koji minimizira potrebnu širinu kanala (podatkovni tok) kojim se to ostvaruje, te koliko iznosi potrebni tok?

Kodiranje izvora radi kompresije

- Odgovor na ova pitanja daje teorija informacije (engl. ***Information Theory, IT***), a posebice njena grana koja se odnosi na kodiranje izvora (engl. ***Source coding***).
- Dio IT-a je teorija koja se bavi izučavanjem odnosa kvalitete i podatkovnog toka i naziva se engl. ***Rate-Distortion theory***.
- Temeljni radovi u ovom području ... Claude Shannon, 1948.-, za vrijeme rada u Bell laboratoriju.

Kodiranje izvora radi kompresije

- Svi moderni koderi medijskih signala intenzivno koriste znanja iz Teorije informacija.
- Ilustrirati ćemo primjenu TI-a na primjeru skalarne kvantizacije s ograničenom entropijom (engl. ***Entropy Constrained Scalar Quantization, ECSQ***).
- ECSQ ima niz praktičnih primjena u kodiranju, a posebice u kodiranju audio signala.
  – ima veliku učinkovitost kod kodiranja dužih nizova statistički nezavisnih (ili barem nekoreliranih) uzoraka.

Skalarna kvantizacija ograničene entropije, ECSQ

- ECSQ kvantizacija se tipično primjenjuje na:
  - uzorke signala u originalnoj vremenskoj ili prostornoj domeni,
  - na koeficijente transformacije signala u nekoj od domena transformacija,
  - na uzorke potpojasnih signala kod struktura koje koriste filtarske slogove za spektralno razlaganje signala,
  - na uzorke signala predikcijske pogreške kod kodera koji koriste prediktore u svrhu vremenske ili prostorne dekorelacije,
  - na paramtere modela, kod kodera temeljenih na modelu sa velikim brojem parametara.

Skalarna kvantizacija ograničene entropije, ECSQ

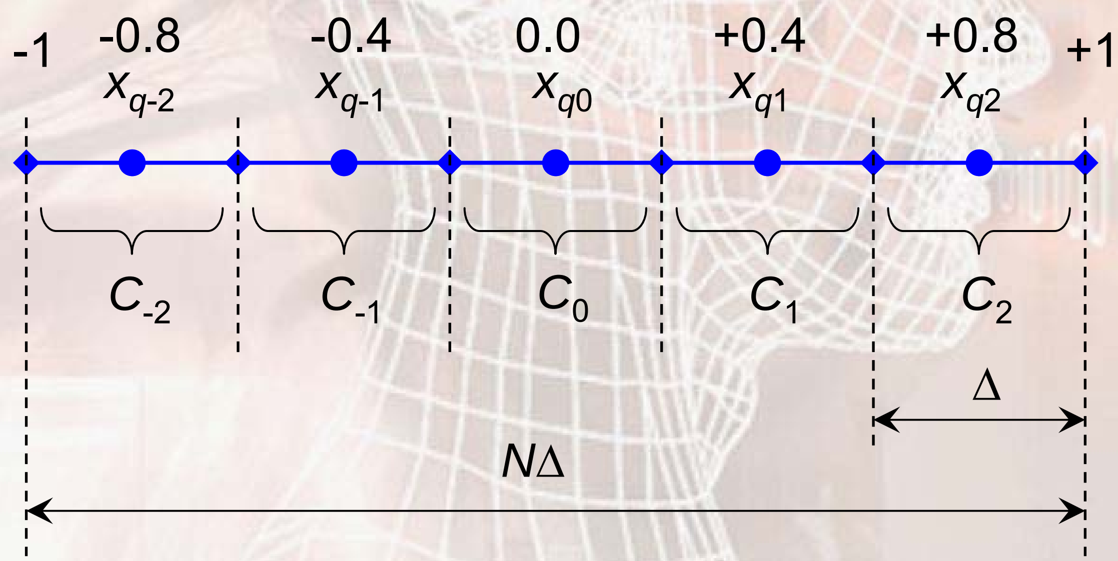
- Ulaz u postupak kodiranja je digitalizirani medijski signal nominalne točnosti;
  - Iako je ovaj signal već amplitudno diskretiziran ulaznim A/D pretvornikom, on se tretira kao “beskonačno” točan, jer je
  - ulazna rezolucija u pravilu mnogo veća od rezolucije ECSQ kvantizatora koji se koristi u sklopu kodera u svrhu kompresije.
  - Obično se koristi bipolarna frakcionalna interpretacija ulaznog signala (+/-1), dijeljenjem cjelobrojne vrijednosti sa 2^{b-1}

Uniformna skalarna kvantizacija-primjer

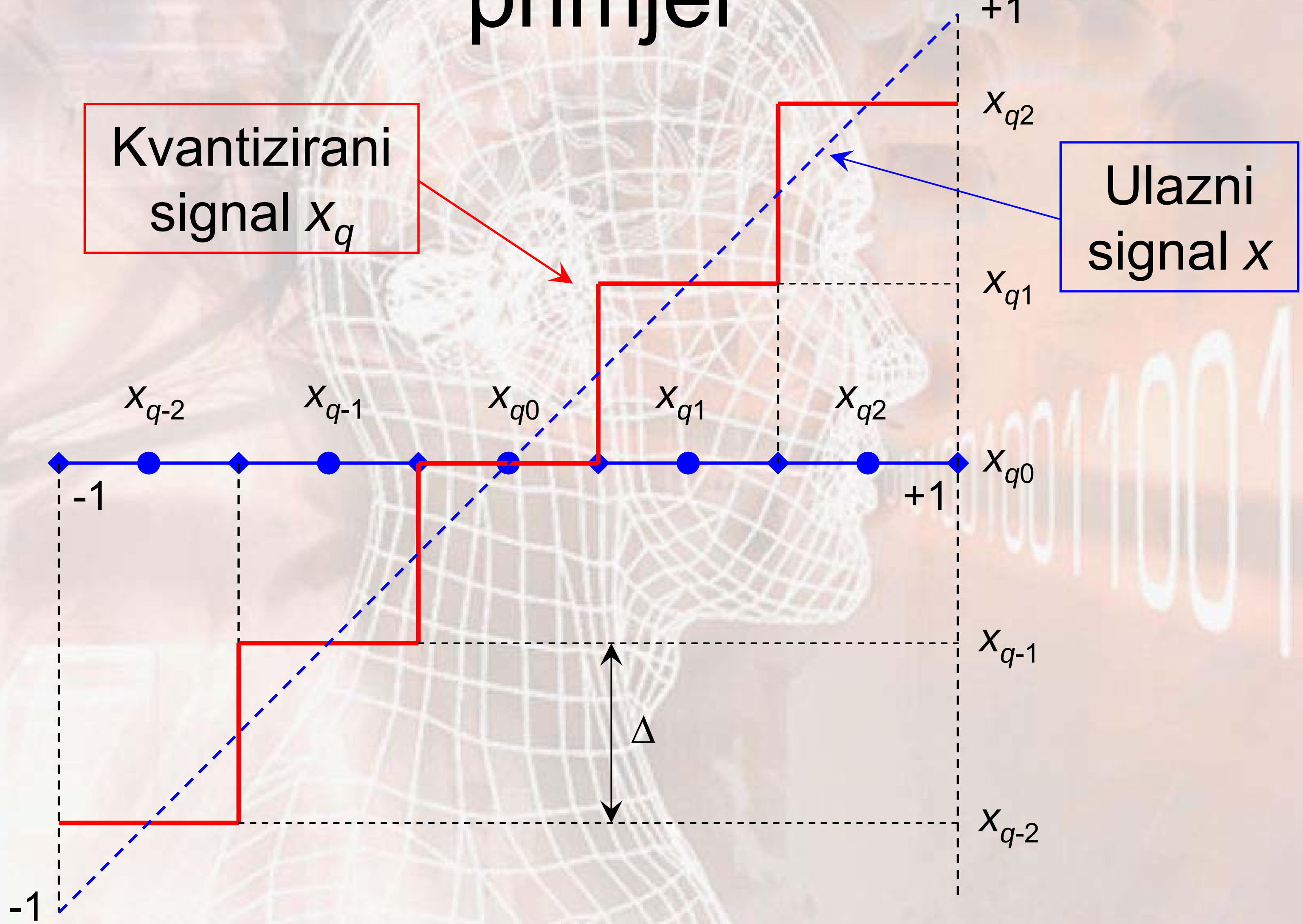
- Primjer:
  - Proces u frakcionalnom zapisu +/- 1 kvantiziramo uniformnim kvantizatorom sa N=5 kvantizacijskih razina, tj. sa korakom $\Delta=2/N=0.4$
  - kvantizacijski razredi su $C_i=[\Delta\cdot i-\Delta/2,\ \Delta\cdot i+\Delta/2)$
    - C_{-2}=[-1.0, -0.6) ...... kod **-2** (centroid x_{q-2}= -0.8)
    - C_{-1}=[-0.6, -0.2) .... kod **-1** (centroid x_{q-1}= -0.4)
    - C_0 =[-0.2, +0.2) ...... kod **0** (centroid x_{q0}= 0)
    - C_1 =[+0.2, +0.6)...... kod **1** (centroid x_{q1}= 0.4)
    - C_2 =[+0.6, +1.0] ..... kod **2** (centroid x_{q2}= 0.8)
  - Rekonstrukciju provodimo **zamjenom koda *i* sa pripadnim centroidom** *i*-tog razreda $x_{qi}=\Delta\cdot i$

Uniformna skalarna kvantizacija-primjer

- ... nastavak:
  - uniformni raspored kvantizacijskih razreda C_i preko područja vrijednosti ulaznog procesa X:



Uniformna skalarna kvantizacija-primjer



Uniformna skalarna kvantizacija-primjer

– Izlazni proces kvantizacijskih kodova *I*={*i* } ima pet simbola $i \in$ {-2, -1, 0, 1 2}
– Kolika je entropija ovog diskretnog izvora *I*?
– ... određena je vjerojatnostima pojedinih simbola:

$$H(I) = -\left(\begin{array}{l} p_I(-2)\log p_I(-2) + p_I(-1)\log p_I(-1) + \\ + p_I(0)\log p_I(0) + p_I(1)\log p_I(1) + p_I(2)\log p_I(2) \end{array} \right)$$

– vjerojatnosti simbola $p_I(i)$ mogu se odrediti integriranjem *pdf* funkcije procesa $f_X(x)$ unutar *i*-tog razreda:

$$p_I(i) = \int_{x \in C_i} f_X(x)dx = \int_{\Delta i - \Delta/2}^{\Delta i + \Delta/2} f_X(x)dx$$

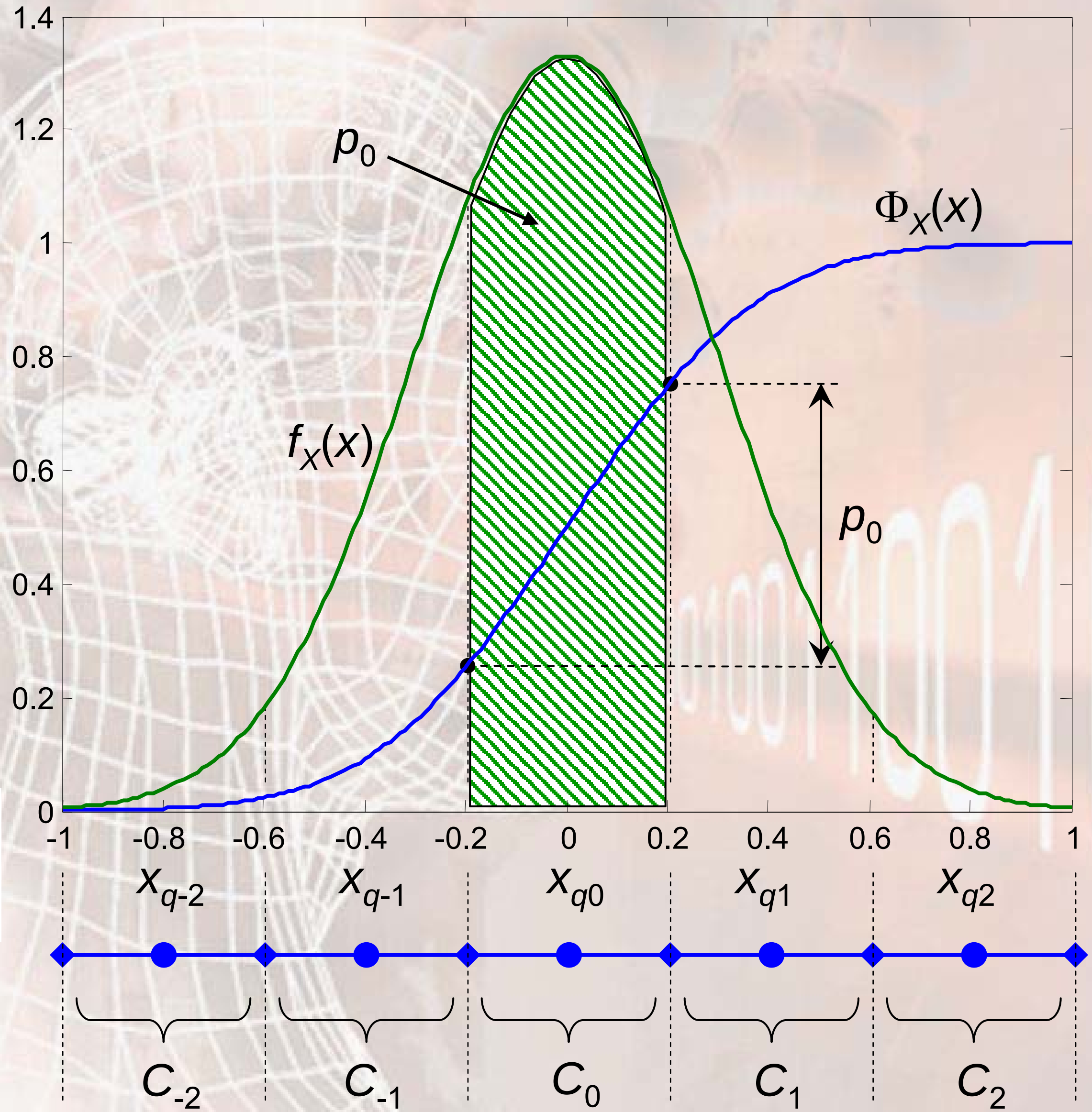
- npr. vjerojatnost simbola *i*=0 je:

$$p_I(0) = \int_{-\Delta/2}^{+\Delta/2} f_X(x)dx$$

- ili diferenciranjem *cdf*-a procesa na rubovima razreda:

$$p_I(0) = \Phi_X(\frac{\Delta}{2}) - \Phi_X(-\frac{\Delta}{2})$$

$$\Phi_X(x) = \int f_X(x)dx$$



Uniformna skalarna kvantizacija-primjer

- ... nastavak:
  - Entropija $H(I)$ je najveća kada su svi izlazni simboli jednako vjerojatni $p_I(i)=1/N$, za i=-2, -1, 0, 1 i 2;
  - To je slučaj kada proces ima uniformnu gustoću:

$$f_X(x) = \frac{1}{\Delta N} \qquad p_I(i) = \frac{1}{\Delta N} \int_{\Delta i - \Delta/2}^{\Delta i + \Delta/2} dx = \frac{1}{\Delta N}\Delta = \frac{1}{N}$$

  - Dobiva se: .... $\max(H(I))=\log(N)=2.32$ bita
  - Napomena: ako koristimo prirodni logaritam u izrazu za entropiju, $H(I)$ se izražava u **natima**, a ako koristimo logaritam baze 2, tada je u **bitovima**!

Uniformna skalarna kvantizacija-primjer

- ... nastavak:
  - Kolika je distorzija *D* uslijed kvantizacije?

$$D = E[(x - x_q)^2] = \int_x f_X(x) \cdot (x - x_q)^2 \, dx \qquad x_q = \Delta \cdot \mathrm{round}(x / \Delta)$$

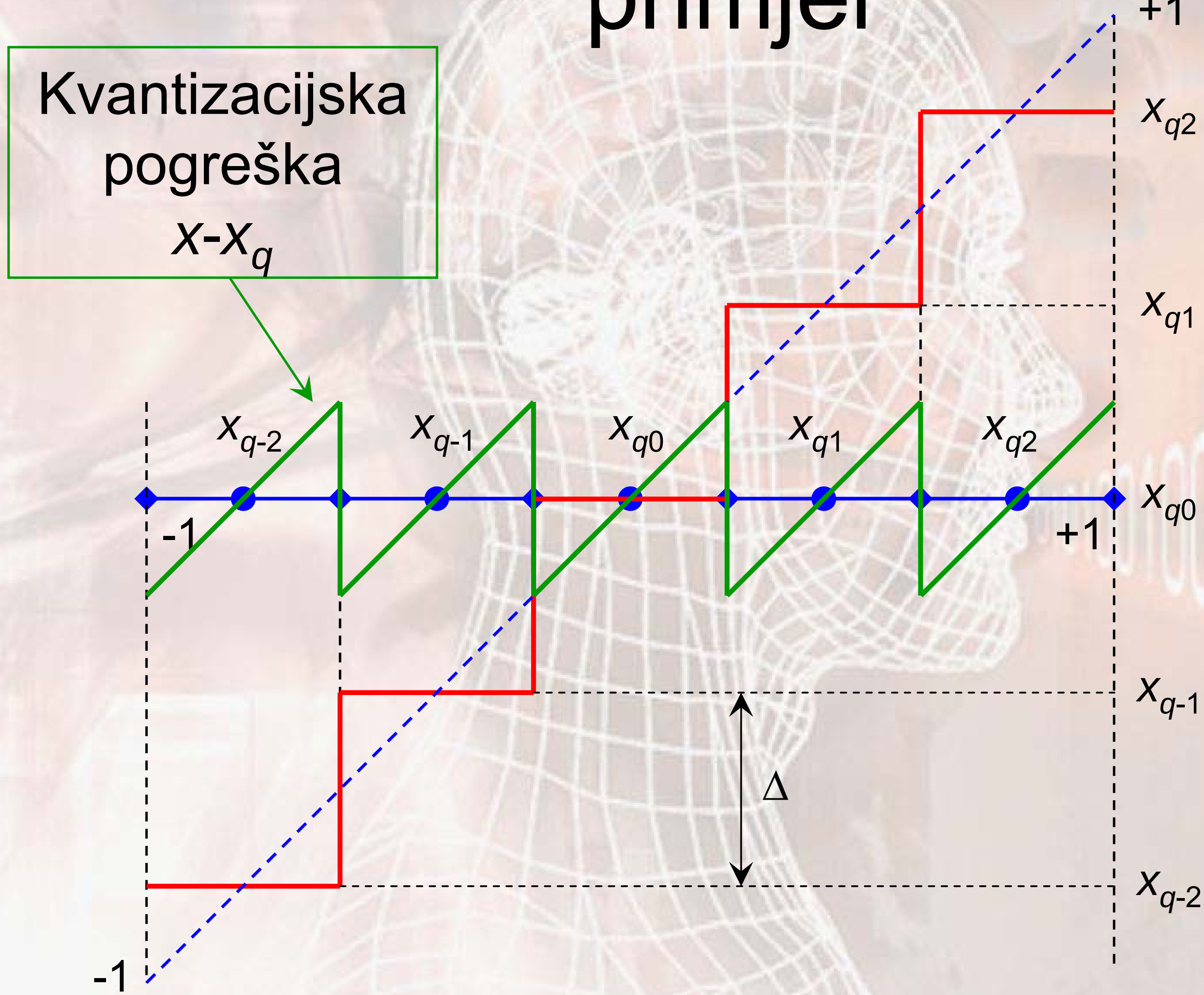
  - Kvantizirani signal x_q je iz konačnog skupa centroida:

$$x_q \in \{x_{q_i}\} = \{x_{q_{-2}}, x_{q_{-1}}, x_{q_0}, x_{q_1}, x_{q_2}\}$$
$$= \{-2\Delta, -\Delta, 0, \Delta, 2\Delta\}$$

  - pa je distorziju *D* je pogodnije prikazati kao sumu ukupne distorzije unutar pojedinih kvantizacijskih razreda D_i

$$D = \sum_i D_i = \sum_i \int_{x \in C_i} f_X(x) \cdot (x - x_{q_i})^2 \, dx$$

Uniformna skalarna kvantizacija-primjer



Uniformna skalarna kvantizacija-primjer

- ... nastavak:
  - Ako je signal uniformne gustoće vjerojatnosti $f_X(x)$=1/(ΔN), distorziju možemo lako odrediti:

$$D=\sum_i \frac{1}{\Delta N}\int_{\Delta i-\Delta/2}^{\Delta i+\Delta/2}(x-\Delta i)^2\,dx=\sum_i \frac{1}{\Delta N}\int_{-\Delta/2}^{+\Delta/2}x'^2\,dx'=\sum_i \frac{\Delta^2}{12N}=\frac{\Delta^2}{12}$$

  - dok varijancu ulaznog procesa nalazimo kao:

$$\sigma_x^2=E[(x-E[x])^2]=E[x^2]=\int_x f_X(x)x^2\,dx$$

$$\sigma_x^2=\frac{1}{\Delta N}\int_{-\Delta N/2}^{\Delta N/2}x^2\,dx=\frac{(\Delta N)^2}{12}$$

Uniformna skalarna kvantizacija-primjer

– Konačno dobivamo *SQNR* odnos za proces uniformne gustoće vjerojatnosti $f_X(x)=1/(\Delta N)$ kvantiziran uniformnim kvantizatorom sa *N* razina koje potpuno pokrivaju područje vrijednosti procesa *X*:

$$SQNR(N) = 10\log_{10}\frac{\sigma_x^2}{D} = 20\log_{10}(N) \quad [dB]$$

– uz entropiju izlaznih simbola *H*(*I*):

$$H(I) = \log_2(N) \quad [bit]$$

– što daje slijedeću jednostavnu vezu:

$$SQNR(N) = (20\log_{10} 2)H(I) = 6.02H(I) \quad [dB]$$

Uniformna skalarna kvantizacija-primjer

- Diskusija:
  - entropija izražena u bitovima i distorzija uslijed kvantizacije izražena u dB su linearno vezane;
  - za smanjenje distorzije od 6dB potrebno je izlaznu entropiju povećati za 1 bit;
  - zbog uniformne razdiobe ulaznog procesa, entropijski koder ne ostvaruje sažimanje varijabilnom dužinom koda, jer su kodovi svih simbola iste dužine;
  - izlazna entropija bilo kojeg drugog ulaznog procesa mora biti manja;
  - distorzija za proizvoljni $f_X(x)$ mora se odrediti zbrajanjem očekivane pogreške D_i koja se nalazi integracijom unutar svakog kvantizacijskog razreda.

Što smo naučili

- kodiranje s ograničenom entropijom / distorzijom
- teorija informacija, grane i primjene
- skalarna kvantizacija s ograničenom entropijom i primjene
- uniformna skalarna kvantizacija – primjer
- odnos kvalitete i entropije

Teorija kvantizacije visokog podatkovnog toka (kvalitete)

prof.dr.sc. Davor Petrinović

Skalarna kvantizacija ograničene entropije, ECSQ

- Problem ...
  - što ako *pdf* funkcija procesa nije uniformna;
  - što ako proces nema ograničeno područje vrijednosti;
  - da li je uopće uniformni raspored centroida kvantizatora i pripadnih razreda najbolje rješenje za proizvoljni *pdf* procesa?
  - kako moraju biti raspoređeni centroidi da minimiziraju pogrešku D za zadanu entropiju $H(I)$, ili obrnuto da minimiziraju entropiju za zadanu distorziju D?

Skalarna kvantizacija ograničene entropije, ECSQ

- Analitičko rješenje nažalost ne postoji za općenit slučaj,
- ... ali se primjenom *teorije kvantizacije visokog podatkovnog toka* (engl. ***High Rate Quantization Theory, HR***) može odrediti asmiptotsko rješenje za visoke rezolucije (odnosno visoku kvalitetu rekonstrukcije).
- Pokazuje se da je ovo rješenje upotrebljivo za ocjenu ponašanja kodera čak i za vrlo male rezolucije!

HR-ECSQ

- Ključna ideja HR teorije je slijedeća:
  - ako su kvantizacijski razredi dovoljno uski, tada se prilikom integracije unutar jednog razreda može pretpostaviti da je funkcija gustoće vjerojatnosti približno konstantna i da se može aproksimirati s vrijednosti *pdf* funkcije u točki centroida tog razreda:

$$f_X(x)\Big|_{x\in C_i} \cong f_X(x_{q_i})$$

  - Ovo značajno olakšava izračunavanje integrala u izrazima za entropiju i distorziju.

HR-ECSQ

- Uz opisanu pretpostavku na visok podatkovni tok dobivamo slijedeće rješenje:
  - ... uistinu, kvantizator s uniformnim razmakom kvantizacijskih nivoa Δ, je najbolje rješenje **za svaki** ulazni proces!
  - Odnos izlazne entropije *H*(*I*) i distorzije *D* uslijed kvantizacije s korakom Δ je:

$$H(I) = h(X) - \frac{1}{2}\log_2(12D) \qquad D = \frac{\Delta^2}{12}$$

  - ... gdje je *h*(*X*) diferencijalna entropija procesa *X*

Diferencijalna entropija

- Diferencijalna entropija predstavlja poopćenje izraza za entropiju na procese kontinuirane varijable ...
  - izraz je potpuno analogan, samo što je suma zamijenjena s integralom preko područja vrijednosti procesa:

$$h(X) = -\int_{X} f_X(x) \cdot \log(f_X(x))\, dx$$

$$h(X) = E[-\log(f_X(x))]$$

HR-ECSQ

- Izraz za odnos entropije i kvalitete ECSQ kvantizatora se može napisati i u slijedećem obliku:

$$SQNR = (20\log_{10} 2)H(I) + SQNR_0 \quad [dB]$$

  – offset ove linearne veze, $SQNR_0$, ovisi o svojstvima ulaznog procesa, ... njegovoj **varijanci i diferencijalnoj entropiji**:

$$SQNR_0 = (20\log_{10} 2)\cdot(\frac{1}{2}\log_2(12\sigma_x^2) - h(X))$$

$$SQNR_0 = 20\log_{10}(\sqrt{12\sigma_x^2}\,2^{-h(X)})$$

HR-ECSQ

- Promotrimo utjecaj člana $SQNR_0$ ...
  - veći offset $SQNR_0$, osigurava bolju kvalitetu za istu izlaznu entropiju $H(I)$;
  - za procese s istom varijancom σ_x^2, manji $h(X)$ implicira veći $SQNR_0$, a time i bolju kvalitetu;
  - Dakle, **diferencijalna entropija *h*(*X*) predstavlja mjeru kompleksnosti ulaznog procesa sa stanovišta učinkovitosti kvantizacije i entropijskog kodiranja**.

HR-ECSQ

- Kako odrediti korak ECSQ kvantizatora za zadanu širinu kanala, tj. ograničenu izlaznu entropiju *H*(*I*)?
  - iz izraza za vezu distorzije i entropije slijedi:

$$\Delta = 2^{h(X)-H(I)}$$

  - Diferencijalnu entropiju *h*(*X*) određujemo iz poznate funkcije gustoće vjerojatnosti ulaznog procesa $f_X(x)$.

HR-ECSQ

- Kako odrediti potrebnu izlaznu entropiju $H(I)_{\mathrm{min}}$ i korak ECSQ kvantizatora Δ za zadanu minimalnu kvalitetu $SQNR_{min}$?
  - Na osnovu $f_X(x)$ računamo $h(X)$ te određujemo offset linearne veze entropije i *SQNRa:*

$$SQNR_0 = 20\log_{10}(\sqrt{12\sigma_x^2}\, 2^{-h(X)})$$

  - Entropiju $H(I)_{\mathrm{min}}$ i korak Δ tada nalazimo kao:

$$H(I)_{\mathrm{min}} = \frac{SQNR_{\mathrm{min}} - SQNR_0}{20\log_{10} 2}$$

$$\Delta = 2^{h(X) - H(I)_{\mathrm{min}}}$$

Diferencijalna entropija – primjer 1

- Odredimo diferencijalnu entropiju za nekoliko tipičnih procesa varijance σ_x^2.
- Proces uniformne gustoće vjerojatnosti:

$$f_{X_{unif}}(x) = \begin{cases} \dfrac{1}{\sqrt{12\sigma_x^2}} & \text{za } |x| < \sqrt{3\sigma_x^2} \\ 0 & \text{inace} \end{cases}$$

$$\Phi_{X_{unif}}(x) = \begin{cases} \dfrac{x + \sqrt{3\sigma_x^2}}{\sqrt{12\sigma_x^2}} & \text{za} |x| < \sqrt{3\sigma_x^2} \\ 0 & \text{inace} \end{cases}$$

Diferencijalna entropija – primjer 1

– Diferencijalna entropija za proces uniformne gustoće vjerojatnosti, nultog očekivanja i zadane varijance σ_x^2 je:

$$h(X_{unif}) = \frac{\log_2(12\sigma_x^2)}{\sqrt{12\sigma_x^2}} \int_{-\sqrt{3\sigma_x^2}}^{\sqrt{3\sigma_x^2}} dx = \log_2\sqrt{12\sigma_x^2}\ [bit]$$

$$SQNR_{0_{unif}} = 10\log_{10}(12\sigma_x^2 \cdot 2^{-2h(X_{unif})})$$

$$= 10\log_{10}(12\sigma_x^2 \cdot 2^{-2\log_2\sqrt{12\sigma_x^2}}) = 0\ [dB]$$

Diferencijalna entropija – primjer 1

- Očito je da veza kvalitete *SQNR* i entropije *H*(*I*) ima nulti offset $SQNR_{0unif}$, što je jednako rezultatu koji smo već prije izveli u primjeru za slučaj procesa uniformne razdiobe.
- Kvantizacijski korak Δ za zadanu izlaznu entropiju kvantizacijskih indeksa *H*(*I*) za proces uniformne razdiobe i varijance σ_x^2 se računa prema slijedećem izrazu:

$$\Delta_{unif} = 2^{h(X_{unif})} \cdot 2^{-H(I)} = \sqrt{12\sigma_x^2} \cdot 2^{-H(I)}$$

Diferencijalna entropija – primjer 2

– Slučajni proces normalne razdiobe s nultim očekivanjem i varijancom σ_x^2, tzv. Gaussov bijeli šum:

$$f_{X_{norm}}(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \cdot e^{-\frac{x^2}{2\sigma_x^2}}$$

$$\Phi_{X_{norm}}(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2\sigma_x^2}}\right)\right)$$

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\,dt$$

– gdje je erf( ) takozvana error funkcija koja je numerički implementirana u Matlabu.

Diferencijalna entropija – primjer 2

– Diferencijalna entropija za slučajni proces normalne razdiobe, nultog očekivanja i varijance σ_x^2 je:

$$h(X_{norm}) = \frac{-1}{\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma_x^2}} \log_2\left(\frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}}\right) dx =$$

$$h(X_{norm}) = \frac{-1}{\ln(2)\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma_x^2}} \left(-\frac{1}{2}\ln(2\pi\sigma_x^2) - \frac{x^2}{2\sigma_x^2}\right) dx =$$

Diferencijalna entropija – primjer 2

– nastavak ... :

$$h(X_{norm}) = \frac{\frac{1}{2}\ln(2\pi\sigma_x^2)}{\ln(2)} \cdot \underbrace{\frac{1}{\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma_x^2}} dx}_{1} + \frac{1}{\ln(2) 2\sigma_x^2} \cdot \frac{1}{\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2\sigma_x^2}} dx$$

$$\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2\sigma_x^2}} dx = \left(-\sigma_x^2 x \cdot e^{-\frac{x^2}{2\sigma_x^2}} + \frac{\sqrt{2\pi\sigma_x^6}}{2} \mathrm{erf}\left(\frac{x}{\sqrt{2\sigma_x^2}}\right) \right) \Bigg|_{-\infty}^{+\infty} =$$

$$= \left(-\sigma_x^2 x \cdot \underbrace{(e^{-\infty} - e^{-\infty})}_{0} + \frac{\sqrt{2\pi\sigma_x^6}}{2} (\underbrace{\mathrm{erf}(\infty)}_{1} - \underbrace{\mathrm{erf}(-\infty)}_{-1}) \right) = \sqrt{2\pi\sigma_x^6}$$

Diferencijalna entropija – primjer 2

– nastavak ... konačno dobivamo $h(X_{norm})$:

$$h(X_{norm}) = \frac{1}{2}\log_2(2\pi\sigma_x^2) + \frac{1}{\ln(2)2\sigma_x^2}\cdot\frac{1}{\sqrt{2\pi\sigma_x^2}}\sqrt{2\pi\sigma_x^6}$$

$$= \frac{1}{2}\log_2(2\pi\sigma_x^2) + \frac{1}{2}\log_2(e) = \frac{1}{2}\log_2(2\pi e\sigma_x^2)\ [bit]$$

$$SQNR_{0_{norm}} = 10\log_{10}(12\sigma_x^2\cdot 2^{-2h(X_{norm})})$$

$$= 10\log_{10}(12\sigma_x^2\cdot 2^{-\log_2 2\pi e\sigma_x^2})$$

$$= 10\log_{10}\left(\frac{12\sigma_x^2}{2\pi e\sigma_x^2}\right) = 10\log_{10}\left(\frac{6}{\pi e}\right) = -1.53\ [dB]$$

Diferencijalna entropija – primjer 2

- Nađimo konačno i izraz za kvantizacijski korak Δ za zadanu entropiju $H(I)$ za šum normalne razdiobe i varijance σ_x^2 :

$$\Delta_{norm} = 2^{h(X_{norm})} \cdot 2^{-H(I)} = \sqrt{2\pi e \sigma_x^2} \cdot 2^{-H(I)}$$

- Zaključno, ... proces normalne razdiobe ima 1.53dB veću distorziju od procesa uniformne razdiobe za istu varijancu i istu izlaznu entropiju.
- To je zato što korak kvantizacije Δ_{norm} mora biti za 1.193 puta veći od Δ_{unif} da bi dobili istu entropiju $H(I)$

$$\frac{\Delta_{norm}}{\Delta_{unif}} = \frac{\sqrt{2\pi e \sigma_x^2}}{\sqrt{12\sigma_x^2}} = \sqrt{\frac{\pi e}{6}} = 1.193$$

HR-ECSQ

- Može se teorijski dokazati da je proces normalne gustoće vjerojatnosti najsloženiji za ECSQ kvantizaciju od svih procesa sa istom varijancom!
- To je zato što ima najveću moguću diferencijalnu entropiju *h*(*X*).
- Dakle, ... za ulazni proces nepoznate *pdf* funkcije kojem poznajemo samo varijancu σ_x^2 uvijek vrijedi:

$$h(X) \le h(X_{norm}) = \log_2 \sqrt{2\pi e \sigma_x^2}\ [bit]$$

- ... što nam može poslužiti kao gornja ograda prilikom projektiranja kvantizatora.

HR-ECSQ

- U okviru primjera 2 pokazali smo:
  - da se skalarni kvantizator sa ograničenom izlaznom entropijom može projektirati i teorijski analizirati čak i za procese sa beskonačnim područjem vrijednosti, što je slučaj za Gaussov proces;
  - u praksi se, zbog projektiranja entropijskog kodera, skup izlaznih simbola ipak ograničava.

Generiranje procesa u Matlabu

- Generator realizacija tri procesa jedinične varijance i nulte srednje vrijednosti:

```
n=[0:N-1];    % vremenska os

w=pi*(0.1+0.8*rand);% slucajna frekvencija za sinus i
                    % trokut ... od 0.1*pi do 0.9*pi
if (sig==1),        % sinusni signal
   y=sin(w*n);      % vrem. diskretni signal
elseif (sig==2),    % slucajni sum normalne razdiobe
   y=randn(1,N);    % generiraj sum
elseif (sig==3),    % pilasti napon
   fa=w*n;          % trenutna faza
   fa=fa-2*pi*floor(fa/2/pi); % osnovna vrijednost faze
   y=fa/pi-1;       % pila od -1 do 1
end;

y=y-mean(y);        % ukloni srednju vrijednost
y=y/sqrt(var(y));   % normaliziraj varijancu na 1
```

Projektiranje ECSQ kvantizatora u Matlabu

- Izračun diferencijalne entropije $h(X)$ i potrebnog kvantizacijskog koraka Δ:

```
% Diferencijalne entropije za nase signale

sigma=1;
if (sig==1),          % za sinus
   hX=1/2*log2(pi^2*sigma^2/2);
elseif (sig==2),     % sum normalne razdiobe
   hX=1/2*log2(2*pi*exp(1)*sigma^2);
elseif (sig==3),     % trokutni signal
   hX=1/2*log2(12*sigma^2);
end;

% Korak kvantizacije za zadanu izlaznu entropiju
% H nalazimo kao:
D=2^(hX-H);
```

Kvantizacija procesa u Matlabu

- Izračun očekivane i stvarne pogreške kvantizacije i sama provedba skalarne kvantizacije:

```
% Ocekivana srednja kvadratna pogreska kvantizacije
% je direknto odredjena sa korakom D
E_er_oc=1/12*D^2;

% Obzirom da su signali jedinicne varijance,
% ocekivani SNR za taj korak kvantizacije je:
SNR_oc(i)=10*log10(1/E_er_oc); % [dB]

% sada provedi stvarnu kvantizaciju realizacije procesa
yi=round(y/D);        % izlazni indeksi kvantizatora
yq=D*yi;              % kvantizirani signal

er=y-yq;              % pogreska kvantizatora
E_er=mean(er.^2);     % ocekivanje kvadrata pogreske
E_y=mean(y.^2);       % ocekivanje kvadrata signala

SNR(i)=10*log10(E_y/E_er); % Stvarni odnos signal sum [dB]
```

Izlazna entropija

- Izračun stvarnog *pdf*-a i entropije izlaznih indeksa kvantizirane realizacije procesa korištenjem histograma:

```
% min & max izlazni kod
yi_min=min(yi);  yi_max=max(yi);

% svi kodovi od min do max
kod=[yi_min:yi_max];

% odredi vjerojatnost pojave pojedinog koda u
% ovoj realizaciji slucajnog procesa koristenjem
% histograma
pdf=hist(yi,kod)/N;

% Odredi entropiju realizacije
kod_postoji=find(pdf>0);
HI(i)=-pdf(kod_postoji)*log2(pdf(kod_postoji))';
```

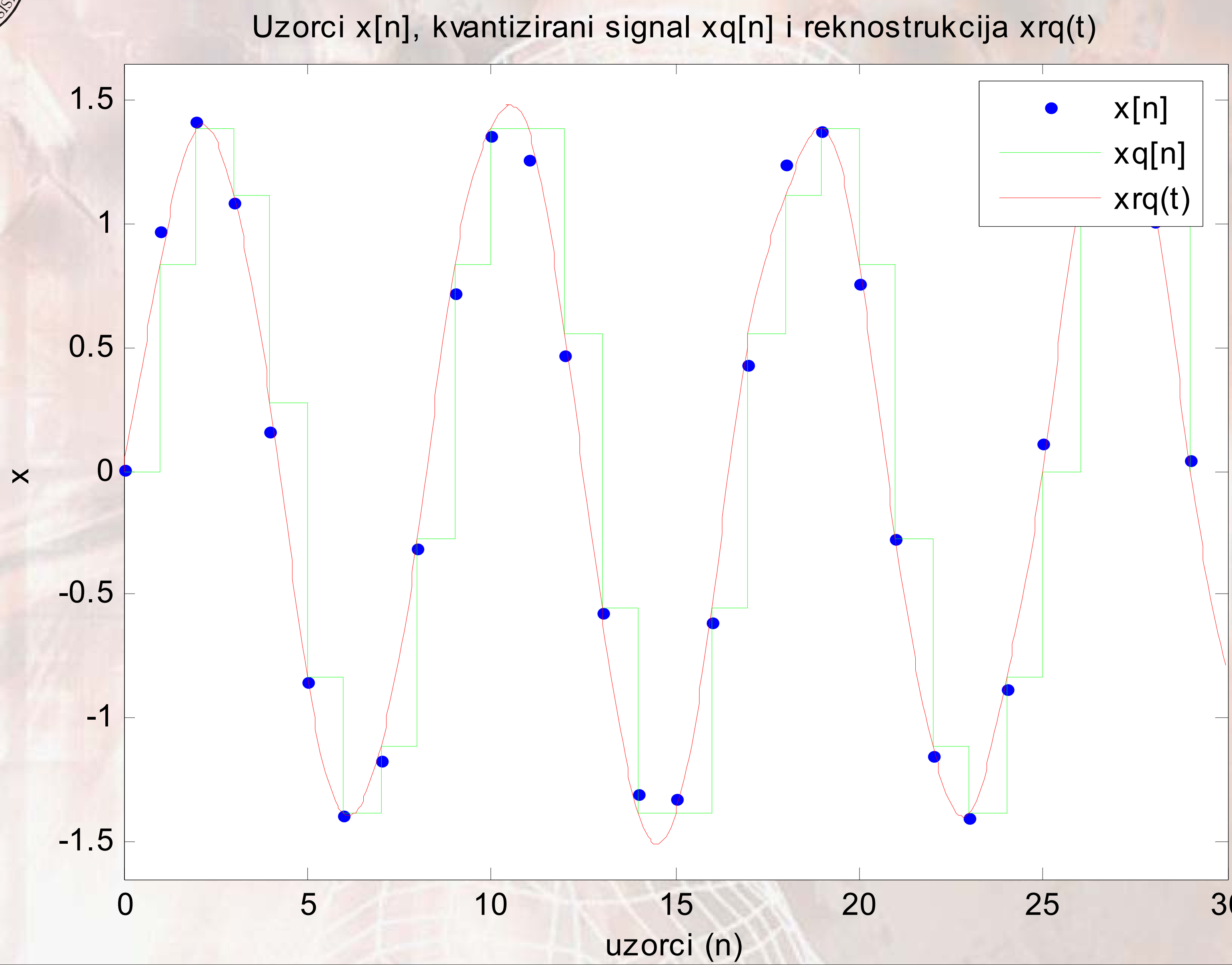
Očekivana izlazna entropija

- Izračun očekivanog *pdf*-a i entropije izlaznih indeksa diferenciranjem analitičkog *cdf*-a ulaznog procesa na rubovima kvantizacijskih razreda:

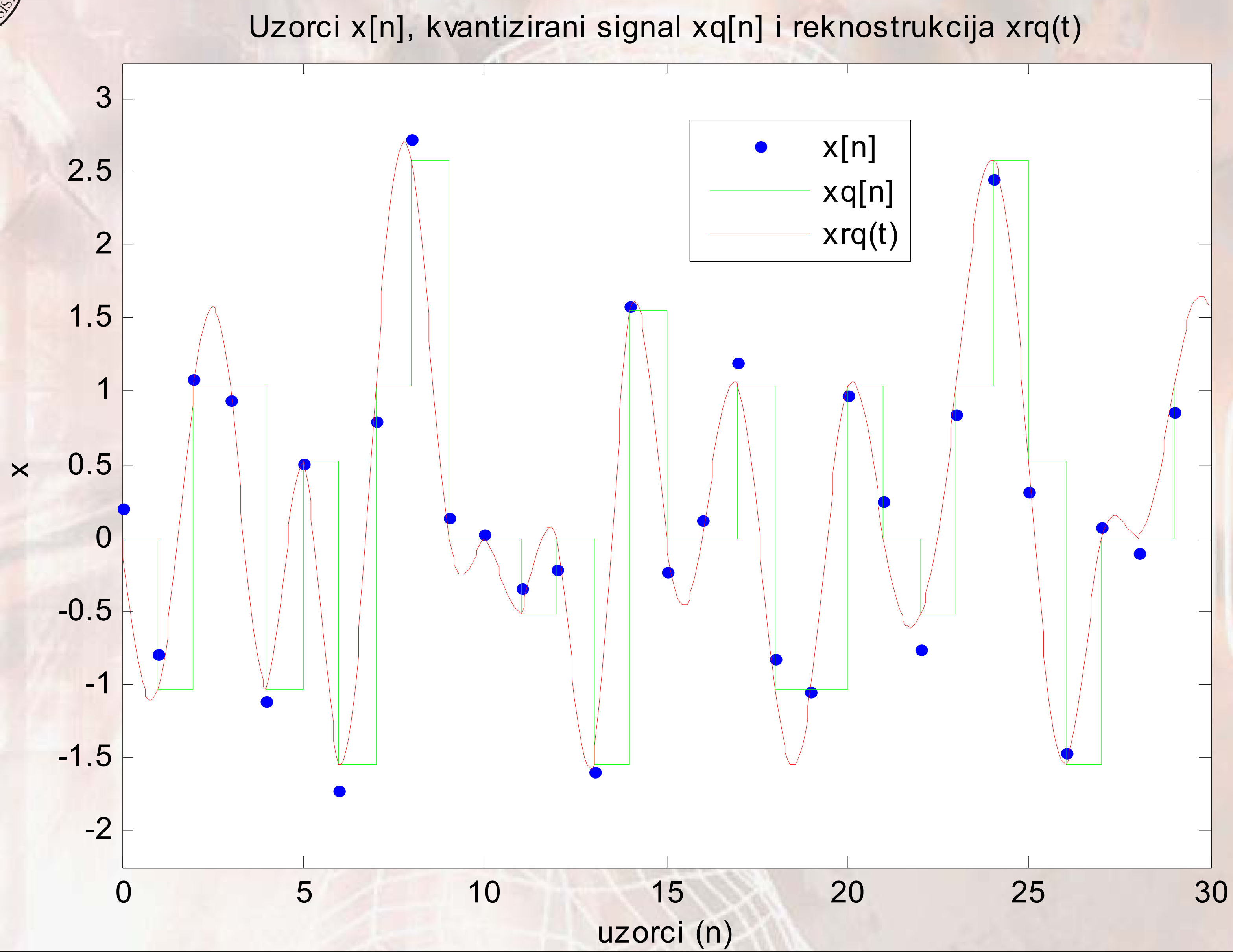
```
delt=max(abs(y));       % tjemeni faktor
% Pragovi kvantizacije u rasponu koristenih kodva
prag=[yi_min-1/2 [yi_min:yi_max]+1/2]*D;

arg=prag/delt;          % argument acos-a i trokuta mora biti
arg=max(min(arg,1),-1); % unutar +/- 1
% Otipkaj cdf funkciju nasih signala jedinicne varijace
if (sig==1),            % za sinusni signal
   cdf_prag=acos(-arg)/pi;
elseif(sig==2),         % za sum normalne razdiobe
   cdf_prag=1/2*(1+erf(prag/sqrt(2)));
elseif(sig==3),         % za pilasti signal
   cdf_prag=(1+arg)/2;
end;
% ocekivana vjerojatnost pojedinog simbola je
% diferencija cdf-a izmedju susjenih pragova
pdf_oc=diff(cdf_prag);
HI_oc(i)=-pdf_oc*log2(pdf_oc)';   % ocekivana izlazna entropija
```

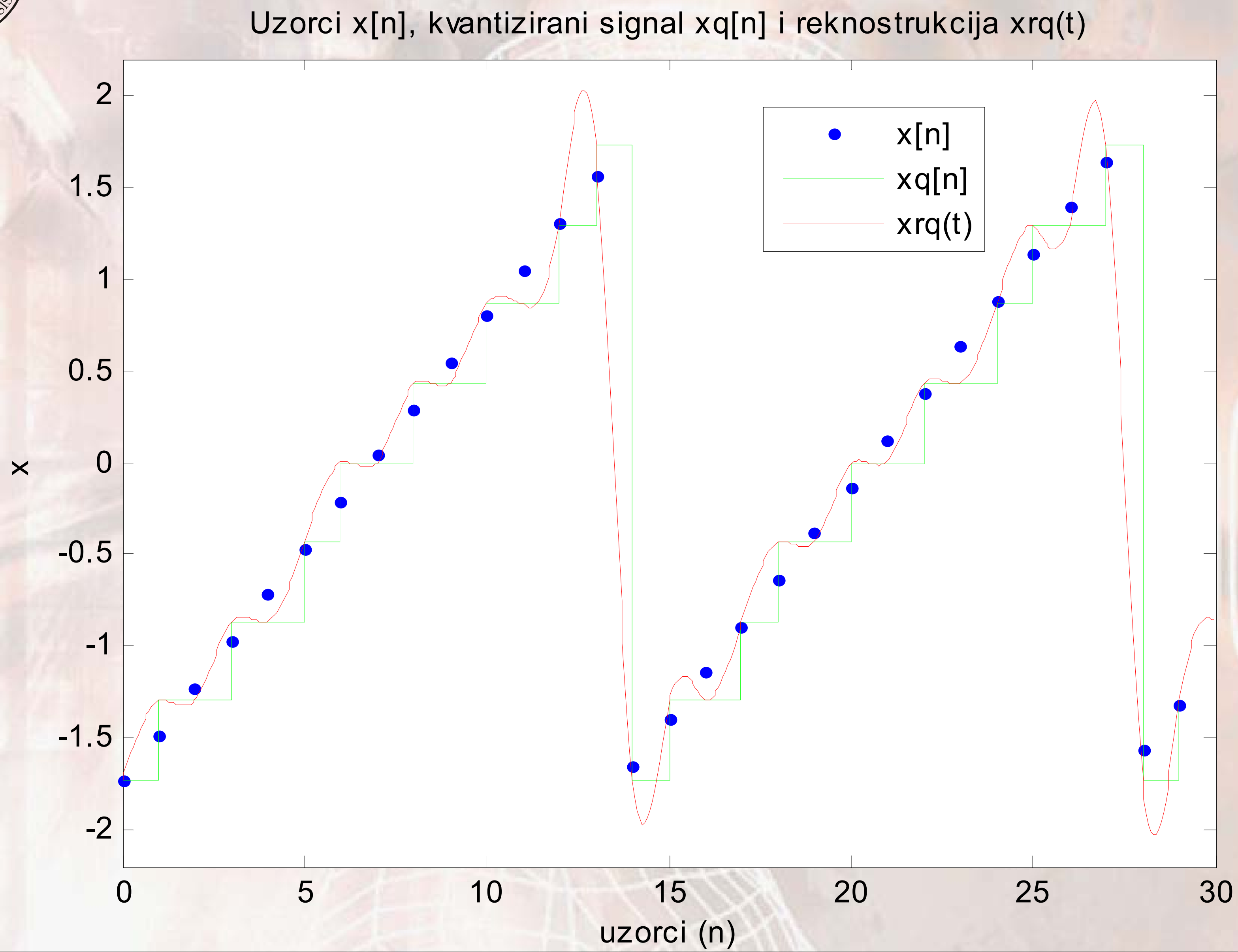
Proces 1 – sinusni signal $\sigma_x^2=1$, $H(I)=3$



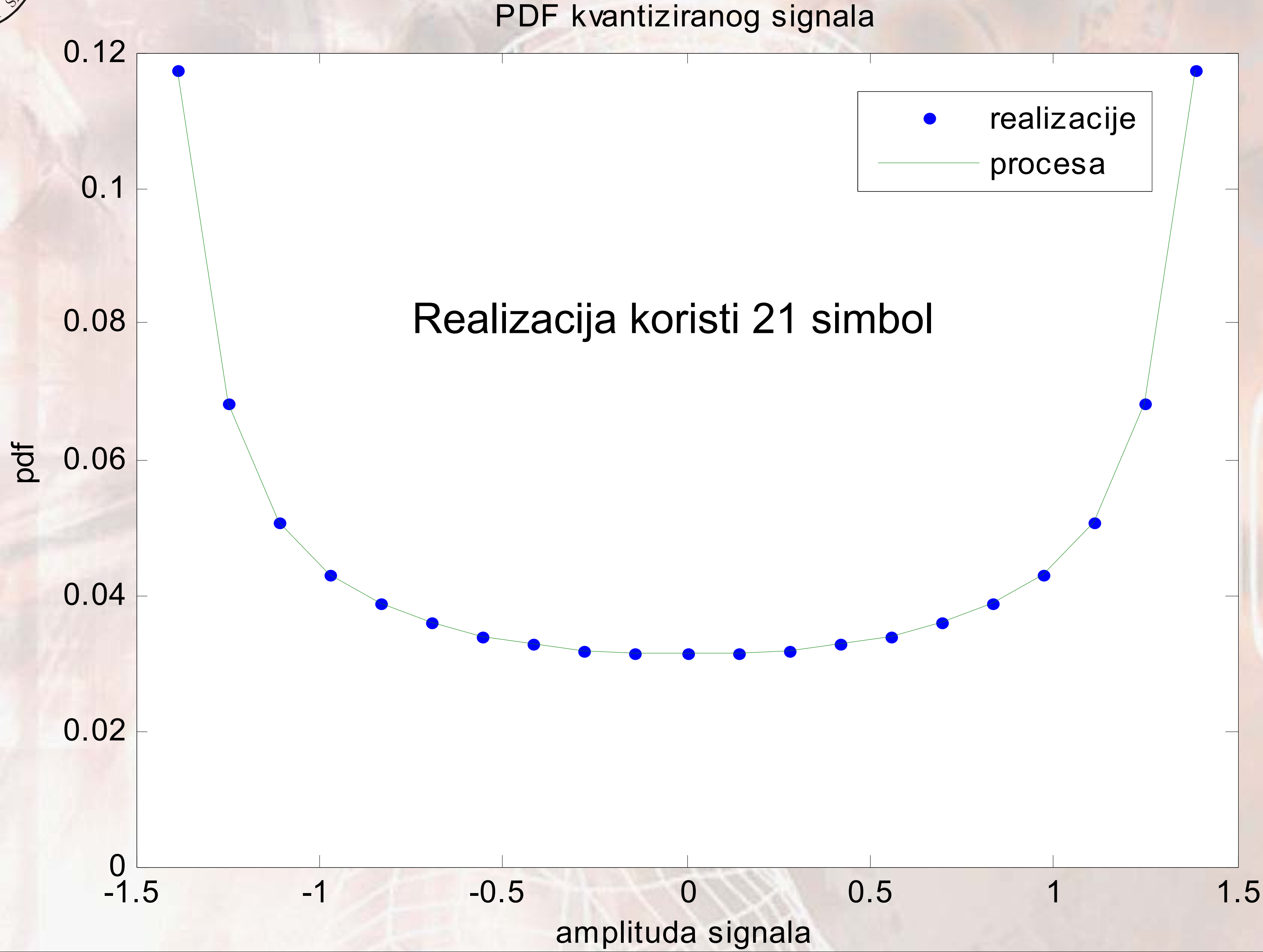
Proces 2 – slučajni šum $\sigma_x^2=1$, $H(I)=3$



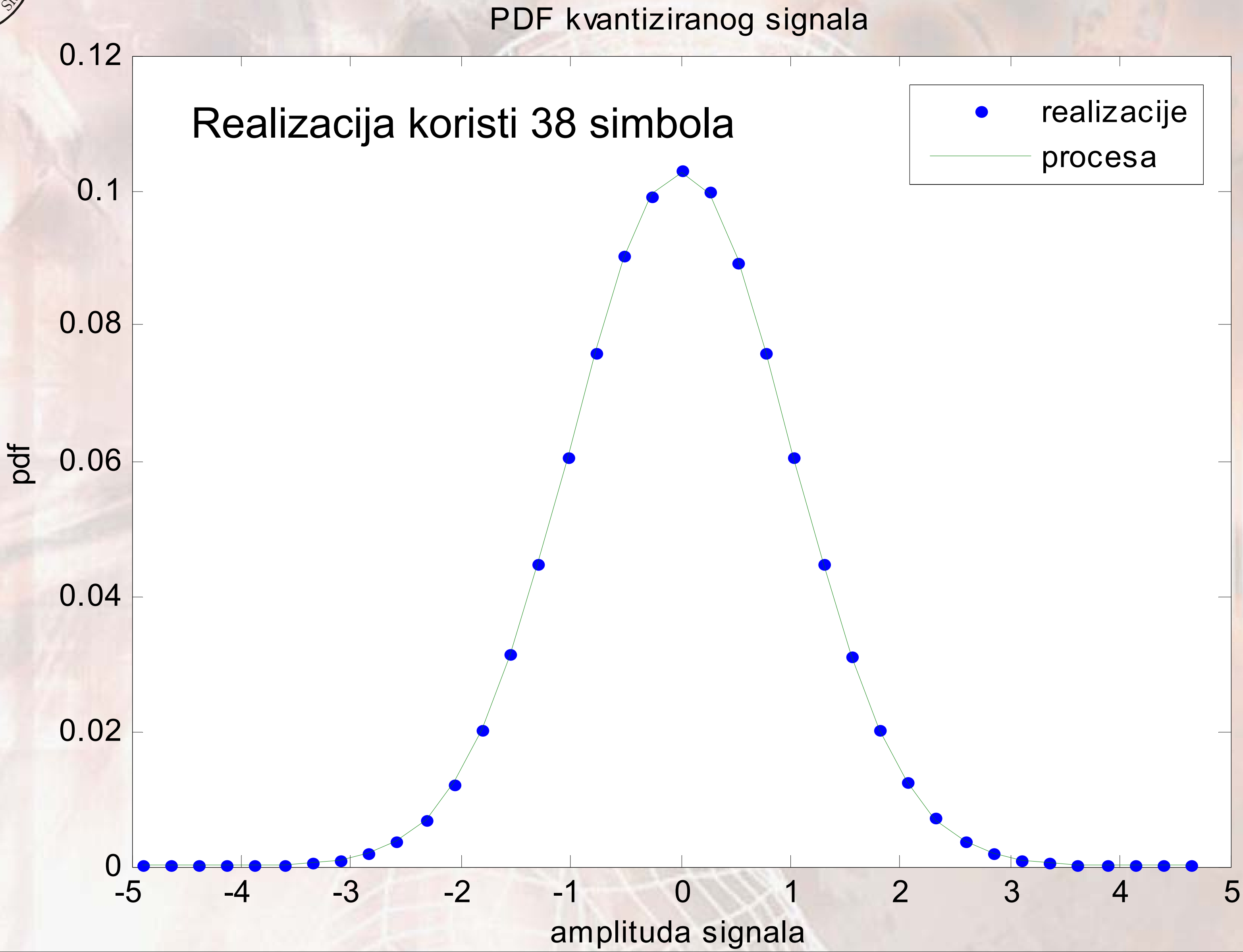
Proces 3 – trokutni signal σ_x^2=1, $H(I)$=3



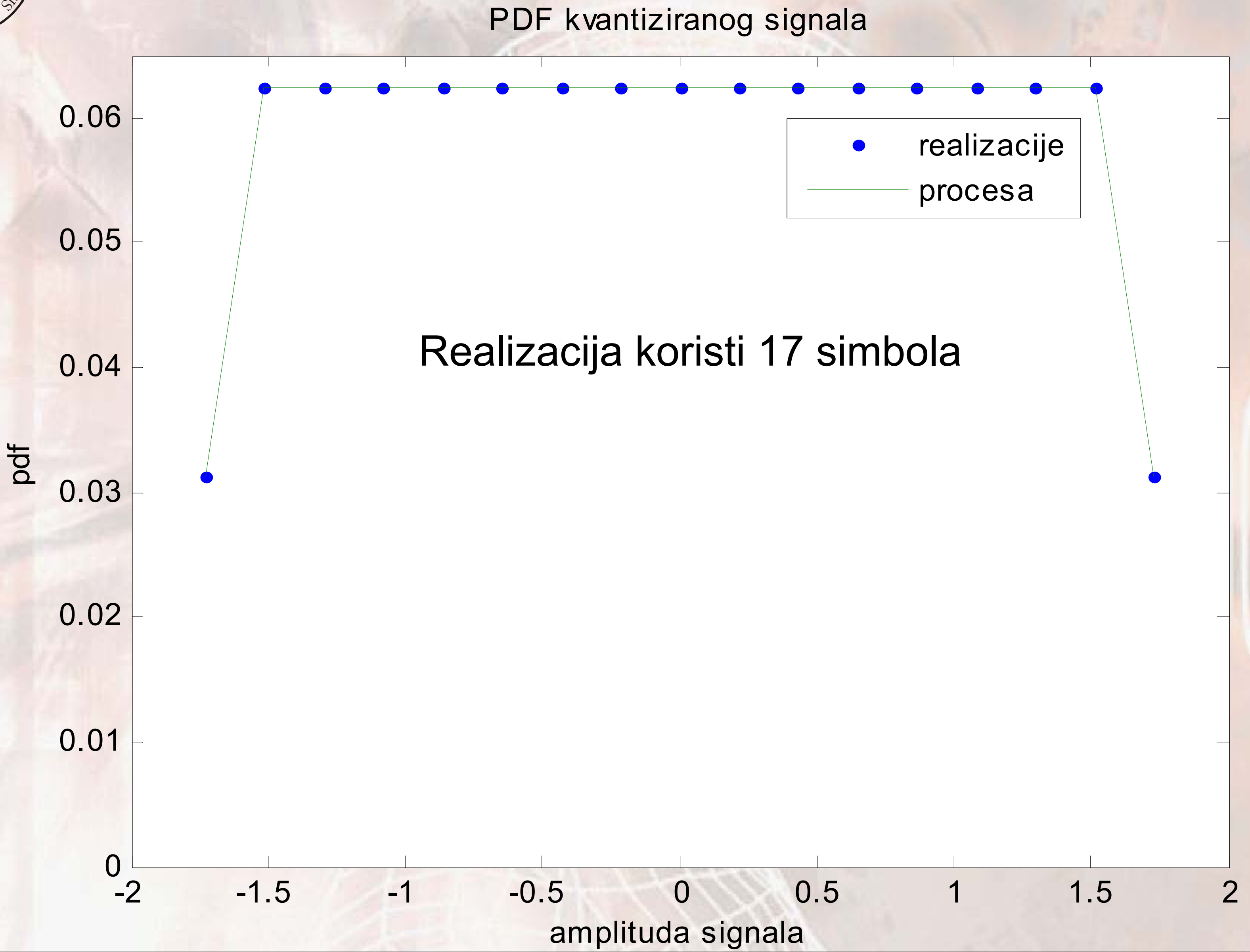
1. Sinus –pdf izlaznih simbola $H(I)$=4



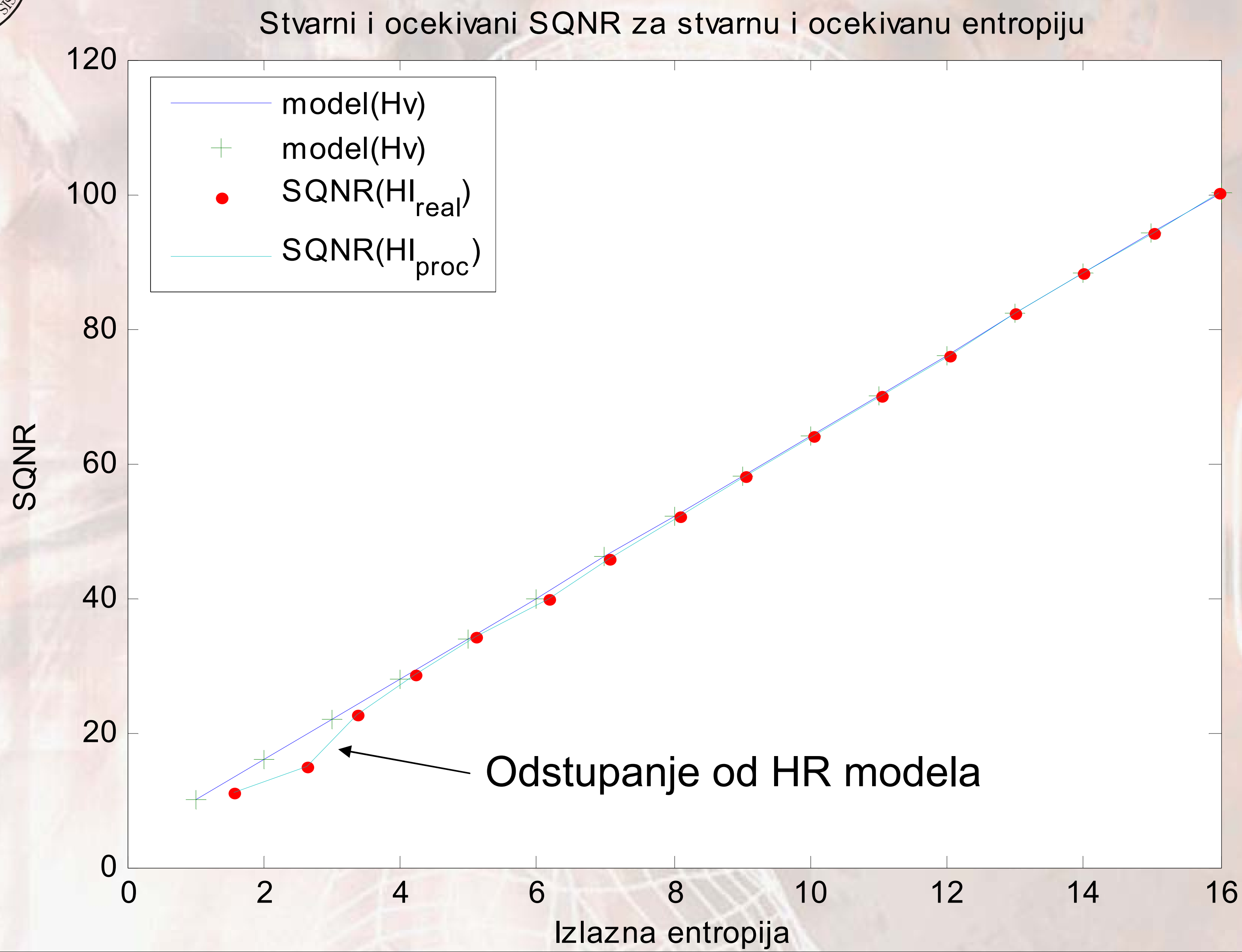
2. Gauss šum –pdf izlaznih simbola *H*(*I*)=4



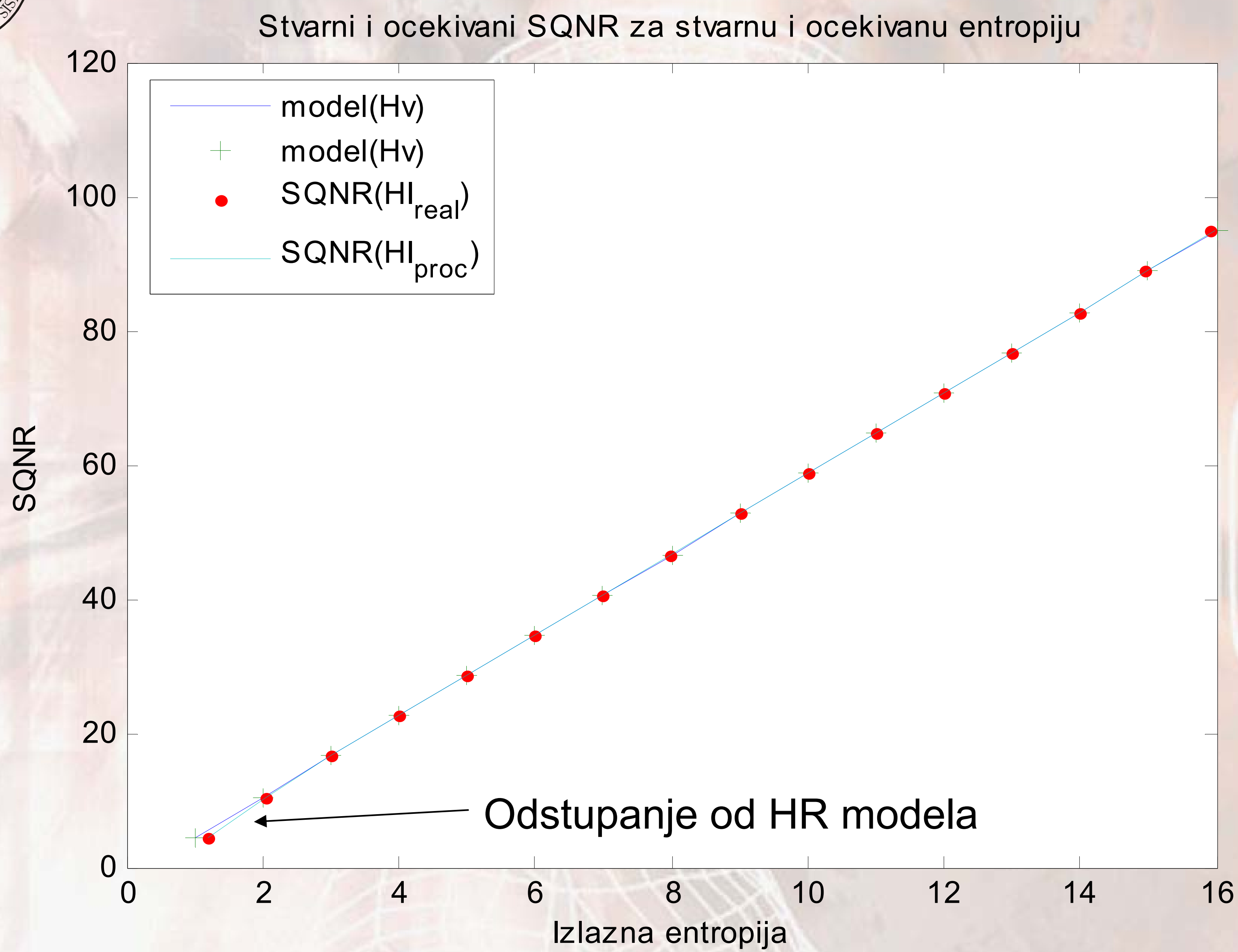
3. Trokut –pdf izlaznih simbola $H(I)$=4



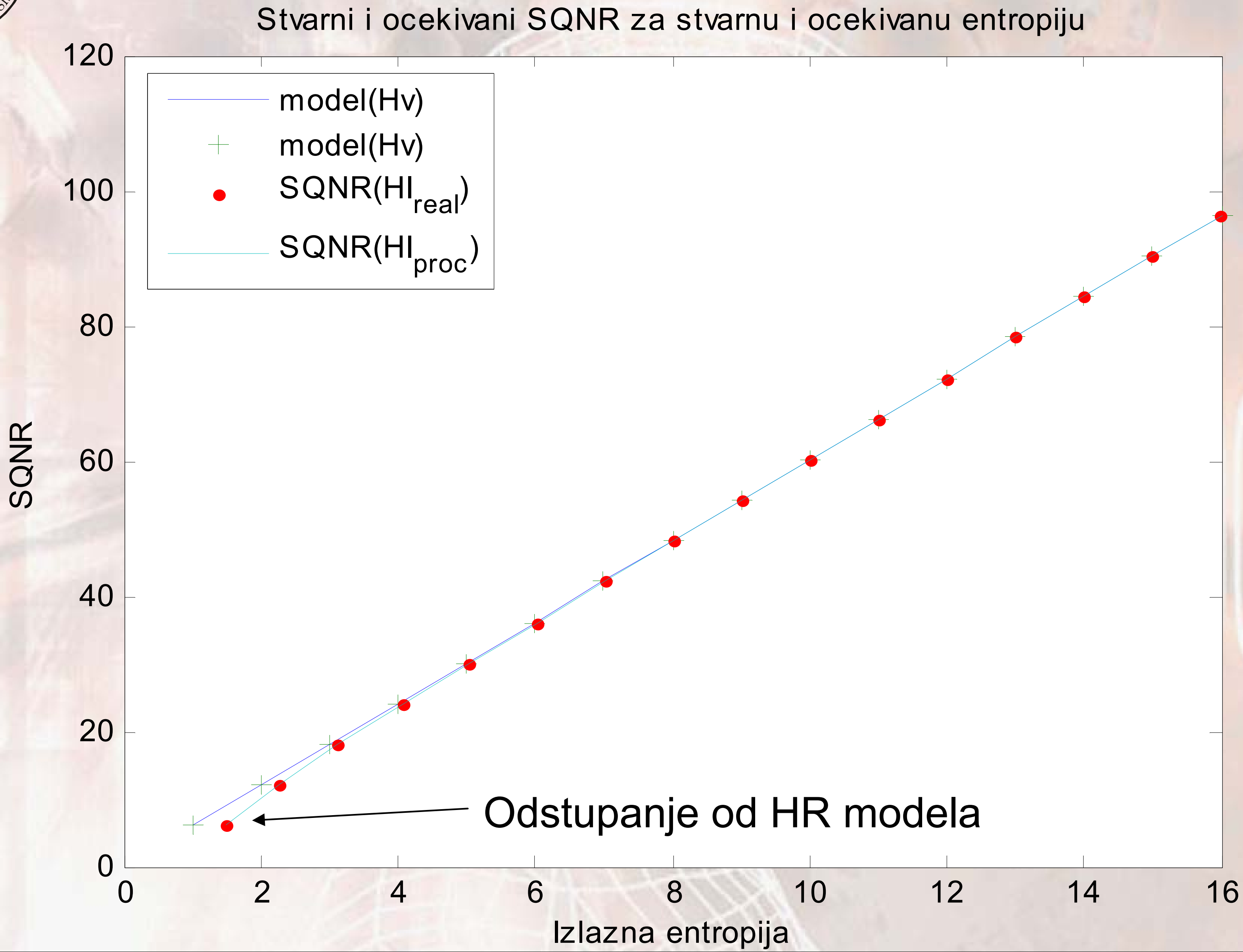
1. Sinus – R-D krivulje $H(I)$=1..16



2. Gauss šum – R-D krivulje *H*(*I*)=1..16



3. Trokut – R-D krivulje $H(I)$=1..16



Što smo naučili

- Teorija kvantizacije visokog podatkovnog toka
- temeljna ideja HR teorije
- odnos entropije i distorzije ECSQ kvantizatora
- diferencijalna entropija
- odnos signal/kvantizacijski šum kao funkcija entropije
- određivanje koraka kvantizatora
- određivanje entropije za zadanu kvalitetu
- primjer procesa uniformne gustoće vjerojatnosti
- primjer slučajnog procesa normalne gustoće vjerojatnosti
- gornja ograda entropije
- primjer HR analize u Matlabu