Slučajni procesi

Regresiona analiza (nastavak). Osnovni pojmovi iz teorije vjerovatnoće. Statistika prvog reda.

Ljubiša Stanković, Miloš Brajović

Univerzitet Crne Gore
Elektrotehnički fakultet

Prezentacija 2

Regresiona analiza. *Ridge regresija*

Ljubiša Stanković, Miloš Brajović

Univerzitet Crne Gore
Elektrotehnički fakultet

Prezentacija 2

Ridge regresija

- Regresija može biti generalizovana na **multivarijabilni slučaj**, kada postoji više od jedne nezavisne promjenljive.
- U ovom slučaju, odbirak signala koji se razmatra, $x(n)$, predstavlja funkciju od slučajnih promjenljivih $t_1(n)$, $t_2(n)$, $\dots$, $t_M(n)$, odnosno, $x(n) = x(t_1(n), t_2(n), \dots, t_M(n))$.
- Regresija može biti zapisana u obliku **višedimenzionog linearnog modela**,

$$x(n) = a_1 t_1(n) + a_2 t_2(n) + \cdots + a_M t_M(n) + \varepsilon(n), \quad n = 1, 2, \dots, N$$

gdje su $t_i(n)$, $i = 1, 2, \dots, M$, nezavisne varijable odbirka $x(n)$.

- Ovaj sistem jednačina može biti zapisan u matrično-vektorskoj formi kao

$$\mathbf{x} = \mathbf{T}\mathbf{a} + \boldsymbol{\Xi}, \quad \text{gdje su:}$$

$$\mathbf{x} = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(N) \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} t_1(1) & t_2(1) & \dots & t_M(1) \\ t_1(2) & t_2(2) & \dots & t_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ t_1(N) & t_2(N) & \dots & t_M(N) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}, \quad \text{i} \quad \boldsymbol{\Xi} = \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \\ \vdots \\ \varepsilon(N) \end{bmatrix}.$$

Ridge regresija

- Uobičajeni cilj je minimizacija kvadrata greške između podataka $\mathbf{x}$ i modela $\mathbf{Ta}$, odnosno:

$$J(\mathbf{a}) = ||\mathbf{x} - \mathbf{Ta}||_2^2 = (\mathbf{x} - \mathbf{Ta})^T(\mathbf{x} - \mathbf{Ta}) = (\mathbf{x}^T - \mathbf{a}^T\mathbf{T}^T)(\mathbf{x} - \mathbf{Ta}).$$

- Rješenje ovog problema minimizacije u smislu najmanjih kvadrata se dobija iz $\partial J(\mathbf{a})/\partial \mathbf{a}^T = \mathbf{0}$, u obliku:

$$-2\mathbf{T}^T(\mathbf{x} - \mathbf{T}\hat{\mathbf{a}}) = \mathbf{0} \quad \text{ili} \quad \mathbf{T}^T\mathbf{x} = \mathbf{T}^T\mathbf{T}\hat{\mathbf{a}}.$$

- Estimacija regresionih parametara $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T$ koji minimizuju funkciju cijene $J(\mathbf{a})$, dobija se iz:

$$\hat{\mathbf{a}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{x} = \mathrm{pinv}\{\mathbf{T}\}\mathbf{x}, \tag{1}$$

  gdje $\mathrm{pinv}\{\mathbf{T}\} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$ predstavlja tzv. **pseudo-inverziju matrice $\mathbf{T}$**.

- Osjetljivost rekonstruisanih koeficijenata na slučajne varijacije u podacima $x(n)$, uzrokovane šumom $\varepsilon(n)$, u velikoj mjeri zavisi od kondicionog broja matrice $\mathbf{T}^T\mathbf{T}$, čija se inverzija vrši.

Ridge regresija

- Po definiciji, **kondicioni broj** proizvoljne normalne matrice $\mathbf{A}$ definiše se kao:

$$\operatorname{cond}\{\mathbf{A}\} = \frac{|\lambda_{\max}|}{|\lambda_{\min}|},$$

  gdje su $\lambda_{\max}$ i $\lambda_{\min}$ najveća i najmanja vrijednost matrice $\mathbf{A}$.
- Matrica $\mathbf{A}$ je normalna ako važi $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H$, gdje je $(\cdot)^H$ Hermitsko transponovanje, odnosno, transponovanje i konjugovanje matrice.
- Za velike vrijednosti kondicionog broja posmatrane matrice $\mathbf{T}^T\mathbf{T}$, koji odgovara relativno maloj vrijednosti determinante $\det\{\mathbf{T}^T\mathbf{T}\}$, mali šum $\varepsilon(n)$ u ulaznim podacima izaziva *veoma velike varijacije* rezultujućih parametara $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_M]^T$ u posmatranom modelu (tzv. ***ill-posed problem***).
- U cilju **regularizacije inverzije** (ali i ograničavanja vrijednosti potencijalno ekstremno velikih elemenata u inverznoj matrici), dodaje se mala vrijednost λ prije inverzije, dok se vektor parametara računa korišćenjem

$$\hat{\mathbf{a}} = (\mathbf{T}^T\mathbf{T} + \lambda\mathbf{I})^{-1}\mathbf{T}^T\mathbf{x}. \qquad (2)$$

- Može se lako pokazati da je ova forma vektora $\mathbf{a}$ rješenje minimizacije funkcije cijene $J(\mathbf{a}) = ||\mathbf{x} - \mathbf{T}\mathbf{a}||_2^2$, kada se za koeficijente $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$, postavi tzv. ograničenje energije.

Ridge regresija

- **Ograničenje energije** u minimizaciji obezbjeđuje da energija vektora $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$ bude najmanja moguća (kako bi se izbjegla pojava velikih vrijednosti u elementima ovog vektora usljed moguće nestabilnosti inverzije matrice $\mathbf{T}^T\mathbf{T}$).
- Ovo je razlog zbog kojeg se regresiona estimacija naziva i *estimacija skupljanja* (engl. ***shrinkage estimation***).
- Ograničena funkcija cijene data je sljedećim izrazom:

$$J(\mathbf{a}) = ||\mathbf{x} - \mathbf{T}\mathbf{a}||_2^2 + \lambda||\mathbf{a}||_2^2,$$

  gdje $||\mathbf{a}||_2^2$ predstavlja L_2-normu vektora $\mathbf{a}$.
- Minimalna vrijednost ove funkcije cijene se dobija iz:

$$\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}^T} = -2\mathbf{T}^T(\mathbf{x} - \mathbf{T}\mathbf{a}) + 2\lambda\mathbf{a} = \mathbf{0}$$

  u obliku datom izrazom (2).
- Ukoliko je ograničenje definisano kao L_1**-norma koeficijenata** $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$, funkcija cijene postaje

$$J(\mathbf{a}) = ||\mathbf{x} - \mathbf{T}\mathbf{a}||_2^2 + \lambda||\mathbf{a}||_1.$$

Ridge regresija

- Za razliku od ograničenja definisanog L_2-normom, koja nameće male vrijednosti (energiju) vektora $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$, ograničenje L_1-normom nameće **rijetko (engl. *sparse*) rješenje** za $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$ (rješenje sa najvećim mogućim brojem elemenata u vektoru $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$ koji su jednaki nuli).
- Minimizacija ove funkcije cijene je poznata kao *operator najmanjeg apsolutnog skupljanja i selekcije* (engl. *Least Absolute Shrinkage and Selection Operator* – **LASSO**). LASSO rješenje se ne može dobiti u zatvorenoj formi.
- Ovaj minimizacioni problem je od ključnog značaja u oblastima kompresivnog odabiranja i mašinskog učenja.

Primjer 4

- Vrši se $N = 10$ mjerenja slučajne varijable $\mathbf{x} = [x(1), x(2), \ldots, x(N)]^T$.
- Poznato je da je slučajna varijabla $x(n)$ linearna funkcija od $M = 7$ nezavisnih promjenljivih $\mathbf{t}(n) = [t_1(n), t_2(n), \ldots, t_M(n)]$.
- Vrijednosti nezavisnih promjenljivih su date u matrici $\mathbf{T}$, gdje $\mathbf{t}(n)$ označava vrste matrice, dok je slučajna varijabla $x(n)$ data u vektoru $\mathbf{x}$.
- Odrediti parametre $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$ linearnog modela korišćenjem *ridge* regresije i $\lambda = 0.01$.

Ridge regresija – primjer 4

Primjer 4 – nastavak

$$\mathbf{T} = \begin{bmatrix} 1.1661 & 1.1607 & 0.5404 & 0.1656 & -0.8154 & -0.7506 & 0.4005 \\ 0.0884 & 0.3245 & 0.4657 & 0.0397 & 1.1129 & -0.1554 & 1.1253 \\ 0.2168 & 0.6992 & -0.2677 & 0.9177 & -0.4409 & -0.5107 & -0.4384 \\ -0.5292 & 0.9924 & 0.8458 & -0.7220 & -0.0592 & 0.9896 & 1.1560 \\ 0.2099 & 0.8470 & -0.5224 & 0.6903 & 0.2075 & 0.1795 & 0.5974 \\ -0.3515 & -1.2374 & 0.1422 & 1.0570 & -0.5864 & 0.2041 & 0.2922 \\ 0.6215 & 1.4229 & 0.5253 & 0.6104 & 0.3214 & -1.1383 & 0.3013 \\ 0.7764 & 0.9572 & 0.5978 & -0.3710 & 1.0122 & -1.2106 & -0.7862 \\ -1.3638 & -0.6115 & -0.1246 & 0.1535 & -0.0534 & 0.6406 & 0.3441 \\ -0.4296 & -0.7374 & 0.6755 & 0.3165 & -0.3022 & 0.4979 & -0.9710 \end{bmatrix} \quad \text{i}$$

$$\mathbf{x} = \begin{bmatrix} 3.1399, & -0.9308, & 0.8620, & -0.9898, & 0.2084, & -0.1228, & 0.9128, & 0.5337, & -2.6691, & -0.5552 \end{bmatrix}^T$$

- Estimirani parametri linearnog modela, koji se dobijaju primjenom *ridge* regresije (2), su:

$$\begin{aligned} \hat{\mathbf{a}} &= (\mathbf{T}^T\mathbf{T} + 0.01\mathbf{I})^{-1}\mathbf{T}^T\mathbf{x} \\ &= [1.9787, -0.0023, -0.0047, -0.0186, -1.0095, -0.0208,\ 0.006]^T. \end{aligned}$$

Primjer 4

- LASSO minimizacija, sa istim $\lambda = 0.01$, daje sljedeći rezultat estimacije:

$$\hat{\mathbf{a}} = \mathrm{lasso}(\mathbf{x}, \mathbf{T}, 0.01) = [1.9767,\ 0,\ 0,\ 0, -0.9776,\ 0,\ 0]^T,$$

  namećući pritom da je najveći mogući broj elemenata u vektoru $\mathbf{a}$ jednak nuli.
- Zbog ovog svojstva, koje je ključno u kontekstu obrade rijetkih signala i kompresivnog odabiranja, LASSO minimizacija može dati rješenje čak i kada je broj opservacija manji od ukupnog broja elemenata u vektoru $\mathbf{a}$.
- Ako zadržimo prvih $N = 6 < M = 7$ vrsta u $\mathbf{T}$ i $\mathbf{x}$, dobijamo

$$\hat{\mathbf{a}} = \mathrm{lasso}(\mathbf{x}, \mathbf{T}, 0.01) = [1.9931,\ 0,\ 0.0004,\ 0, -0.9772,\ 0,\ 0]^T.$$

- U ovom slučaju važi $\det\{\mathbf{T}^T\mathbf{T}\} = 0$, budući da je $N < M$.
- LASSO minimizaciona procedura za L_1-normu je implementirana u Octave/MATLAB-u.
- U našim razmatranjima, navedeni optimizacioni pristup ćemo koristiti ne zalazeći u njegovu implementaciju.

Polinomijalna aproksimacija

- Specifična forma linearne regresije je **polinomijalna aproksimacija** (engl. ***polynomial fitting***):

$$x(t_n) = a_0 + a_1 t_n + a_2 t_n^2 + \cdots + a_M t_n^M + \varepsilon(t_n), \quad n = 1, 2, \ldots, N.$$

- Ovakav oblik regresije je i dalje linearan, zato što linearnost važi u odnosu na parametre modela $a_0, a_1, \ldots, a_M$.
- Matrica nezavisnih varijabli je data u sljedećem obliku:

$$\mathbf{T} = \begin{bmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^M \\ 1 & t_2 & t_2^2 & \ldots & t_2^M \\ \vdots & & & & \\ 1 & t_N & t_N^2 & \ldots & t_N^M \end{bmatrix},$$

  pri čemu su svi drugi vektori, rezultati i komentari identični kao u multivarijabilnom slučaju.
- U kontekstu polinomijalnih aproksimacija, regularizaciona ograničenja nad rješenjem sprečavaju ,,preprilagođavanje“ (engl. ***overfitting***) modela i održavaju niske vrijednosti parametara.

Metod kros-validacije

- U praksi, određivanje najbolje vrijednosti parametra λ predstavlja složen problem. Jedan pristup određivanju najbolje vrijednosti λ je poznat pod nazivom **metod kros-validacije**.
- Navedeni metod se sastoji od sljedećih koraka:
  1. Estimacija parametara modela se radi sa unaprijed odabranim skupom vrijednosti λ, gdje se jedna ili više tačaka $(t_n, x(n))$ *isključuje* iz razmatranja.
  2. Estimacija se ponavlja, ali uz isključivanje ostalih tačaka, jedne po jedne.
  3. Računa se ukupna kvadratna greška vrijednosti predviđenih modelom u odnosu na izostavljene vrijednosti, za svaku razmatranu vrijednost λ.
  4. Najbolja vrijednost λ je ona koja daje najmanju ukupnu kvadratnu grešku predikcije.

Osnovni pojmovi iz teorije vjerovatnoće. Statistika prvog reda.

Ljubiša Stanković, Miloš Brajović

Univerzitet Crne Gore
Elektrotehnički fakultet

Prezentacija 2

Vjerovatnoća

- Teorija vjerovatnoće je naučna disciplina koja se bavi analizom slučajnih fenomena na bazi skupa aksioma.
- Vjerovatnoća je veličina kojom se kvantifikuje u kolikoj mjeri je ishod određenog slučajnog događaja vjerovatan, odnosno, vjerovatnoća predstavlja kvantifikaciju očekivanja da će se neki događaj desiti.
- Za računanje parametara *statistike prvog reda*, dovoljno je kao osnovnu probabilističku deskripciju poznavati **vjerovatnoću** ili **funkciju gustine vjerovatnoće** slučajne promjenljive.
- Pretpostavimo da slučajni signal $x(n)$ može uzeti jednu od diskretnih vrijednosti (amplituda) ξ_i iz skupa $\mathbb{A} = \{\xi_1, \xi_2, \ldots, \xi_N\}$.
- U posmatranom kontekstu, govorimo o vjerovatnoćama da slučajni signal $x(n)$ u trenutku n uzima specifičnu vrijednost ξ_i iz skupa svih mogućih vrijednosti

$$\mathrm{P}\{x(n) = \xi_i\} = P_{x(n)}(\xi_i).$$

gdje operator $\mathrm{P}\{\cdot\}$ označava vjerovatnoću odgovarajućeg događaja.

Vjerovatnoća

- Funkcija vjerovatnoće $P_{x(n)}(\xi)$ zadovoljava sljedeće uslove (*aksiome* teorije vjerovatnoće):
  1. $0 \leq P_{x(n)}(\xi) \leq 1$ za bilo koje ξ.
  2. Za događaje $x(n) = \xi_i$ i $x(n) = \xi_j$, $i \neq j$, koji su međusobno **isključivi** važi:
  $$\mathrm{P}\left\{x(n) = \xi_i \text{ ili } x(n) = \xi_j\right\} = P_{x(n)}(\xi_i) + P_{x(n)}(\xi_j).$$
  3. **Suma vjerovatnoća** da $x(n)$ uzima bilo koju vrijednost ξ_i iz skupa $\mathbb{A}$ svih mogućih vrijednosti ξ je **siguran događaj**. Vjerovatnoća sigurnog događaja je jednaka 1, pa važi:
  $$\sum_{\xi \in \mathbb{A}} P_{x(n)}(\xi) = 1.$$
- **Nemoguć događaj** ima vjerovatnoću jednaku nuli.
- Primjer signala kod kojeg su vjerovatnoće estimirane nakon eksperimenta (*a posteriori*) već je prezentovan u sklopu ranijih predavanja.
- ***A posteriori* vjerovatnoća** da signal $x(n)$ uzima vrijednost ξ_i je definisana kao odnos između broja pojavljivanja N_{ξ_i} događaja $x(n) = \xi_i$ i ukupnog broja sprovedenih eksperimenata N:
$$P_{x(n)}(\xi_i) = \frac{N_{\xi_i}}{N}.$$
za dovoljno veliko N i N_{ξ_i}.

Vjerovatnoća

- U nekim slučajevima, moguće je odrediti vjerovatnoću događaja **prije** nego što se eksperiment sprovede.
- Na primjer, ako je signal jednak broju koji se pojavljuje prilikom bacanja kocke, (gdje je pretpostavljeno da su sve strane kocke ravnopravne), tada taj signal može uzeti jednu od vrijednosti iz skupa $\xi_i \in \mathbb{A} = \{1,2,3,4,5,6\}$.
- U ovom slučaju, vjerovatnoća svakog događaja je poznata unaprijed (**a priori**), i jednaka je $P(\xi_i) = 1/6$.
- **Nezavisnost.** Dva događaja (dva slučajna odbirka signala) su nezavisni jedan od drugog, *ukoliko vjerovatnoća da se jedan događaj desi* (da jedan odbirak uzme slučajnu vrijednost) *ne utiče na vjerovatnoću da se desi drugi događaj* (ne utiče na vrijednost drugog odbirka signala). Ako su signali $x(n)$ i $x(m)$ **statistički nezavisne slučajne varijable**, tada

$$\mathrm{P}\left\{x(n) = \xi_i \text{ i } x(m) = \xi_j\right\} = P_{x(n)}(\xi_i) P_{x(m)}(\xi_j).$$

- **Isključivost.** Dva slučajna događaja su međusobno isključiva *ukoliko se ne mogu desiti u isto vrijeme*. Ako su vrijednosti signala $x(n) = \xi_i$ i $x(n) = \xi_j$ međusobno isključivi događaji, tada, prema osobini (2) važi:

$$\mathrm{P}\left\{x(n) = \xi_i \text{ ili } x(n) = \xi_j\right\} = P_{x(n)}(\xi_i) + P_{x(n)}(\xi_j).$$

Vjerovatnoća – primjer

Primjer 1

Razmatra se slučajni signal čije su vrijednosti jednake broju koji se pojavljuje pri bacanju kocke. Skup mogućih vrijednosti signala je $\xi_i \in \mathbb{A} = \{1,2,3,4,5,6\}$.

- Odrediti vjerovatnoću da odbirak signala uzme vrijednost $x(n) = 2$ ili vrijednost $x(n) = 5$, odnosno $\mathrm{P}\{x(n) = 2 \text{ ili } x(n) = 5\}$.
- Odrediti vjerovatnoću da odbirak signala u trenutku n uzme vrijednost $x(n) = 2$ i da pri sljedećem bacanju signal uzme vrijednost $x(n+1) = 5$, odnosno

$$\mathrm{P}\{x(n) = 2 \text{ i } x(n+1) = 5\}.$$

- Događaji da $x(n) = 2$ i $x(n) = 5$ su međusobno isključivi. Stoga, vjerovatnoća dva isključiva događaja jednaka je sumi njihovih pojedinačnih vjerovatnoća,

$$\mathrm{P}\{x(n) = 2 \text{ ili } x(n) = 5\} = P_{x(n)}(2) + P_{x(n)}(5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

- Događaji da $x(n) = 2$ i $x(n+1) = 5$ predstavljaju statistički nezavisne događaje. U ovom slučaju važi:

$$\mathrm{P}\{x(n) = 2 \text{ i } x(n+1) = 5\} = P_{x(n)}(2)P_{x(n)}(5) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}.$$

Uslovna vjerovatnoća

- Uslovna vjerovatnoća je vjerovatnoća da se događaj A desi, pod uslovom da se drugi događaj B već desio. Vjerovatnoća događaja A, pod uslovom da se desio događaj B se kratko zapisuje u formi $P(A|B)$.
- Vjerovatnoća da su se desila oba događaja A i B je:

$$\mathrm{P}\{A \text{ i } B\} = P(A|B)P(B),$$

gdje je $P(B)$ vjerovatnoća da se desio događaj B, dok $P(A|B)$ označava vjerovatnoću da se događaj A desio pod uslovom da se događaj B već desio.

Primjer 2

Pretpostaviti da je dužina slučajnog signala $x(n)$ jednaka N, a da je I odbiraka oštećeno ekstremno jakim šumom. Neka se kao skup mjerenja (opservacija) uzima podskup od $M < N$ slučajno pozicioniranih odbiraka signala.

- Koja je vjerovatnoća da u skupu od M slučajno pozicioniranih odbiraka signala nema odbiraka koji su oštećeni jakim šumom?
- Ako je $N = 128$, $I = 16$, i $M = 32$, odrediti koliko se skupova od M odbiraka, koji ne sadrže nijedan odbirak koji je oštećen jakim šumom, može očekivati u 1000 realizacija (pokušaja).

Primjer 2 – rješenje

- Vjerovatnoća da prvi slučajno odabrani odbirak nije oštećen jakim šumom može biti izračunata kao *a priori* vjerovatnoća,
$$P(1) = P(B) = \frac{N-I}{N}$$
budući da postoji N odbiraka i da je njih $N-I$ neoštećeno (od strane šuma).
- Nakon što se odabere prvi odbirak bez šuma, u preostalih $N-1$ odbiraka, postoji tačno $N-1-I$ odbiraka koji su neoštećeni šumom. Vjerovatnoća da se odabere odbirak bez šuma je sada $P(A|B) = (N-1-I)/(N-1)$.
- Vjerovatnoća da drugi slučajno odabrani odbirak nije oštećen jakim šumom, pod uslovom da prvi odabrani odbirak nije oštećen, jednaka je proizvodu vjerovatnoća
$$P(2) = P(A) = P(A|B)P(B) = \frac{N-1-I}{N-1}\,\frac{N-I}{N}.$$
Ovdje se koristi svojstvo uslovne vjerovatnoće.
- Dalje se nastavlja proces slučajne selekcije odbiraka. Na isti način se može odrediti vjerovatnoća da svih M slučajno odabranih odbiraka nijesu pod uticajem jakog šuma kao:
$$P(M) = \frac{N-I}{N}\,\frac{N-1-I}{N-1}\cdots\frac{N-(M-1)-I}{N-(M-1)} = \prod_{i=0}^{M-1}\frac{N-I-i}{N-i}.$$
- Za $N = 128$ odbiraka signala, sa $I = 16$ odbiraka oštećenih šumom, vjerovatnoća da $M = 32$ slučajno selektovanih odbiraka nije oštećeno šumom jednaka je $P(32) = 0.0071$.

Primjer 2 – rješenje

- Ako se ista procedura sa selekcijom $M = 32$ odbirka ponovi 1000 puta, možemo očekivati
$$P(32) \times 1000 = 7.1,$$
što je oko 7 realizacija u kojima nijedan od M odbiraka signala nije zahvaćeno šumom. Jedna realizacija bez oštećenja šumom se očekuje na svakih 140 realizacija.
- U literaturi, može se naići na sljedeći **proračun očekivanog broja iteracija** neophodnih da bi se došlo do realizacije bez šuma.
- Vjerovatnoća da je jedan odbirak bez šuma (*inlier*) jednaka je $(N-I)/N$. Zatim se pretpostavlja da ova vjerovatnoća može da se iskoristi za M odbiraka (odbirak se vraća i može se opet birati).
- Vjerovatnoća da je najmanje jedan odbirak oštećen jakim šumom u ukupno M odbiraka je $[1-((N-I)/N))^M]$.
- Najzad, vjerovatnoća da se dobije realizacija bez jakog šuma u ukupno N_{it} takvih pokušaja je
$$P = 1-[1-((N-I)/N))^M]^{N_{it}} \quad \text{gdje je} \quad N_{it} = \frac{\ln(1-P)}{\ln(1-((N-I)/N))^M)}.$$
- Navedeni proračun je tačan ako je $(N-I-M)/(N-M) \sim (N-I)/N$, inače se umjesto $((N-I)/N))^M$ mora koristiti
$$P(M) = \prod_{i=0}^{M-1} \frac{N-I-i}{N-i}.$$

Bajesova teorema

- Razmatraju se dva događaja A i B. **Bajesova relacija** se dobija iz vjerovatnoće da se dese oba događaja:

$$\mathrm{P}\{A \text{ i } B\} = P(A|B)P(B) = \mathrm{P}\{B \text{ i } A\} = P(B|A)P(A).$$

kao:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

- Pretpostavimo sada da se posmatra N mogućih događaja A_i, $i = 1, 2, \ldots, N$. Tada važi:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}.$$

- Pretpostavimo dalje da su događaji A_i, $i = 1, 2, \ldots, N$ **nezavisni** i **iscrpni**.
- Navedeno znači da dva događaja A_i i A_j, $i \neq j$, ne mogu da se dese istovremeno (nezavisnost) i da jedan od događaja A_i, $i = 1, 2, \ldots, N$, mora da se desi (iscrpnost). U tom slučaju, može se pisati

$$P(B) = \mathrm{P}\{B \text{ i (siguran događaj)}\} = \mathrm{P}\{B \text{ i } (A_1 \text{ ili } A_2 \text{ ili } \ldots A_N)\}$$
$$= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_N)P(A_N).$$

Bajesova teorema

- **Bajesova teorema** je data izrazom:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_N)P(A_N)}.$$

- Pošto je Bajesovski pristup od velikog značaja u savremenoj obradi podataka, u nastavku će mu biti posvećeno više detalja.

1. Događaj B je **pretpostavljen** (**dokaz** se desio). Vjerovatnoća $P(B)$ pokazuje koliko je vjerovatan pretpostavljeni događaj (dokaz) pod uslovom svih mogućih događaja (hipoteza) A_i. Navedeno se može zapisati u obliku

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_N)P(A_N)$$

2. Specifični događaj A_i se naziva **hipoteza**, pretpostavljajući da se događaj B desio. Vjerovatnoća $P(A_i)$ pokazuje koliko je vjerovatan događaj A_i (hipoteza), nezavisno od toga da li se događaj B desio.
3. Vjerovatnoća $P(B|A_i)$ pokazuje koliko je vjerovatan događaj B (dokaz), pod uslovom da se događaj A_i desio (tj. pod uslovom da je hipoteza tačna).
4. Rezultat $P(A_i|B)$ je vjerovatnoća događaja A_i (koliko je vjerovatna hipoteza A_i) uz činjenicu da se dokaz B desio.

Bajesova teorema – primjer

Primjer 3

Razmatraju se četiri slike, označene sa A_1, A_2, A_3, i A_4. U dvije slike (A_1 i A_2) postoji 20% crvenih piksela, u trećoj slici (A_3) postoji 30% crvenih piksela, dok u četvrtoj slici (A_4) postoji 50% crvenih piksela. Bira se slučajno jedna slika i posmatra se jedan njen piksel. Odabrani piksel je crvene boje (dokaz B). Koja je vjerovatnoća da je odabrana slika A_4?

- Vjerovatnoća da je slika A_4 odabrana (hipoteza A_4) kada se crveni piksel dobija kao dokaz (označeno kao B) jednaka je:

$$P(A_4|B) = \frac{P(B|A_4)P(A_4)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+P(B|A_3)P(A_3)+P(B|A_4)P(A_4)}.$$

- Vjerovatnoća da je dobijen crveni piksel, u oznaci $P(B)$, pod uslovom da su se desili svi mogući događaji (hipoteze) A_i, $i = 1,2,3,4$, data je izrazom:

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+P(B|A_3)P(A_3)+P(B|A_4)P(A_4) \\ &= \frac{20}{100}\frac{1}{4}+\frac{20}{100}\frac{1}{4}+\frac{30}{100}\frac{1}{4}+\frac{50}{100}\frac{1}{4} = 0.3. \end{aligned}$$

Primjer 3 – rješenje

- Ova vjerovatnoća je dobijena kao suma po svim događajima A_i, i poznata je kao **marginalna vjerovatnoća**.
- Vjerovatnoća da se desila hipoteza A_4, nezavisno od boje piksela (nezavisno od događaja B) je
$$P(A_4) = 1/4 = 0.25,$$
budući da postoje četiri slike, čiji je izbor jednako vjerovatan, $P(A_1) = P(A_2) = P(A_3) = P(A_4) = 1/4$.
- Vjerovatnoća $P(B|A_4)$ govori koliko je vjerovatan događaj B (dokaz u vidu crvenog piksela), pod uslovom da se događaj A_4 desio (hipoteza da je četvrta slika već odabrana). Njena vrijednost je
$$P(B|A_4) = \frac{50}{100} = 0.5,$$
budući da je 50% crvenih piskela u slici A_4.
- Rezultujuća vjerovatnoća da je slika A_4 odabrana, pod uslovom da se događaj B desio (crveni piksel se pojavio) je jednaka:
$$P(A_4|B) = \frac{0.5 \times 0.25}{0.3} = 0.4167.$$

Bajesova teorema – primjer

Primjer 4

Razvijen je sistem za testiranje virusa, za koji je poznata očekivana pouzdanost. Vjerovatnoća korektnih pozitivnih ili negativnih rezultata za testiranu osobu je veoma visoka, i jednaka je 0.978. Vjerovatnoća pogrešno pozitivnog (engl. *false-positive*) rezultata (testirana osoba nema virus, ali je test pozitivan) iznosi $P_{F+} = 0.018$. Vjerovatnoća pogrešno negativnog (engl. *false-negative*) rezultata (testirana osoba ima specificirani virus, ali je rezultat negativan) je $P_{F-} = 0.004$.

- Koliki je očekivani broj pozitivnih rezultata na 1000 testiranih osoba iz zemlje u kojoj je očekivani broj osoba zaraženih virusom:
  1. 1 na 1000 osoba ($p = P(V) = 10^{-3}$);
  2. 1 na 10,000 osoba ($p = P(V) = 10^{-4}$);
  3. 1.75 na 100 osoba ($p = P(V) = 0.0175$); i
  4. 25 na 100 osoba iz odabranog skupa za testiranje (koji je formiran na bazi prethodne evaluacije drugih simptoma)?
- Slučajno odabrana osoba se testirala na virus i rezultat je pozitivan. Odrediti vjerovatnoću da je ova osoba zaražena virusom u sva četiri prethodna slučaja.

Primjer 4 – rješenje

- Vjerovatnoća pozitivnog rezultata je jednaka sumi vjerovatnoće da testirana osoba nema virus i da je rezultat pozitivan, $(1-P(V))P_{F+}$, i vjerovatnoće da osoba ima virus i da je rezultat pozitivan, $P(V)(1-P_{F-})$.
- Navedeno znači da je test slučano odabrane osobe pozitivan sa vjerovatnoćom

$$P(+) = (1-P(V))P_{F+} + P(V)(1-P_{F-}) = (1-p)P_{F+} + p(1-P_{F-}).$$

- Za dati očekivani broj osoba zaraženih virusom, $p = P(V)$, dobijamo:
  1. $P(+) = 0.019$, što znači da će biti 10 pozitivnih rezultata na 1000 testiranih osoba, iako je očekivani odnos 1 na 1000, u ovom slučaju. Stoga, većina rezultata testa će biti pogrešno pozitivna.
  2. $P(+) = 0.0181$ potvrđuje da pogrešno pozitivni rezultati ponovo dominiraju u testu.
  3. U ovom slučaju, dobija se $P(+) = 0.035$, što znači da polovina pozitivnih rezultata odgovara osobama koje su kontaminirane virusom, budući da bi ova vjerovatnoća, sa idealnim testom, značila da je 3.5 od 100 osoba kontaminirano.
  4. Za odabranu grupu osoba sa simptomima, sa relativno visokom vjerovatnoćom da su zaraženi virusom, dobija se vjerovatnoća $P(+) = 0.2625$, što znači da je poklapanje sa očekivanim brojem kontaminiranih osoba visoko.
- Zaključak je da nasumično testiranje (engl. *random screening*) populacije ne može dati zadovoljavajuće rezultate ukoliko je vjerovatnoća pojave virusa mala a test nije idealan (ako test ne daje nula pogrešno pozitivnih i negativnih ishoda).

Primjer 4 – rješenje

- Testiranje treba raditi na prethodno odabrani skup ljudi. Ovaj zaključak je još jasniji u kontekstu Bajesove analize, date u nastavku.
- Kada se slučajno odabrana osoba testira, i rezultat je pozitivan, tada je *a posteriori* vjerovatnoća događaja da je osoba kontaminirana virusom, ukoliko je test pozitivan, $P(V|+)$, jednaka:

$$P(V|+)=\frac{P(+|V)P(V)}{P(+)}=\frac{(1-P_{F-})P(V)}{(1-P(V))P_{F+}+P(V)(1-P_{F-})}=\frac{(1-P_{F-})p}{(1-p)P_{F+}+p(1-P_{F-})},$$

  gdje je $P(+|V)$ vjerovatnoća pozitivnog testa, ako je osoba kontaminirana virusom, $p = P(V)$ je vjerovatnoća da slučajna osoba ima virus, dok je $P(+)$ vjerovatnoća da je rezultat testa pozitivan, uključujući oba slučaja da osoba ima i da osoba nema virus.
- Za tri razmatrana slučaja, vrijednosti vjerovatnoće $P(V|+)$ su: (a) $P(V|+) = 0.0525$, (b) $P(V|+) = 0.0055$, i (c) $P(V|+) = 0.4964$.
- (d) Za odabrani skup osoba sa značajnom vjerovatnoćom da su kontaminirani virusom, dobija se $P(V|+) = 0.9486$, što znači *visoku pouzdanost testa* (ako je test pozitivan osoba je kontaminirana).
- Ovi rezultati potvrđuju zaključaka da *slučajno testiranje osoba ne treba raditi*, osim ukoliko je vjerovatnoća pojave virusa kod testiranih osoba povećana usljed drugih simptoma, što znači da će skup testiranih osoba imati virus sa značajnom vjerovatnoćom.

Srednja vrijednost i varijansa

- Vidjeli smo da se **srednja (prosječna) vrijednost** računa za skup dostupnih uzoraka eksperimenta, te da je kao takva **slučajna po svojoj prirodi**.
- Ukoliko je dostupan probabilistički opis slučajne varijable (signala), tada je moguće predvidjeti srednju (prosječnu) vrijednost signala *bez korišćenja njegovih specifičnih realizacija*, odnosno, bez eksperimenata sa slučajnim pokušajima.
- Ovakva, **analitički dobijena**, srednja vrijednost, poznata je pod nazivom **očekivana vrijednost** i predstavlja pravu srednju vrijednost, koja bi se dobila na osnovu velikog broja eksperimenata. Ona je **deterministička veličina**.

- **Očekivana vrijednost** signala $x(n)$ se računa kao suma po skupu dostupnih amplituda, $\xi \in \mathbb{A} = \{\xi_1, \xi_2, \ldots, \xi_N\}$, sa odgovarajućim vjerovatnoćama kao težinskim koeficijentima te sume, odnosno:

$$\mu_x(n) = \mathrm{E}\{x(n)\} = \sum_{\xi \in \mathbb{A}} \xi P_{x(n)}(\xi).$$

- **Varijansa** odbirka slučajnog signala $x(n)$ koja uzima vrijednosti ξ iz slučajnog skupa $\mathbb{A} = \{\xi_1, \xi_2, \ldots, \xi_N\}$, sa poznatim vjerovatnoćama, $P_{x(n)}(\xi_i)$, definiše se kao

$$\sigma_x^2(n) = \mathrm{E}\{|x(n) - \mu_x(n)|^2\} = \sum_{\xi \in \mathbb{A}} |\xi - \mu_x(n)|^2 P_{x(n)}(\xi).$$

Srednja vrijednost i varijansa – primjer

Primjer 5

Slučajni signal $x(n)$ može uzeti vrijednosti iz skupa $\xi \in \mathbb{A} = \{0, 1, 2, 3, 4, 5\}$. Poznato je da je za $k = 0, 1, 2, 3, 4$ vjerovatnoća da je $x(n) = k$ dva puta veća od vjerovatnoće da je $x(n) = k + 1$. Odrediti vjerovatnoće $P_{x(n)}(\xi_k) = P\{x(n) = k\}$. Odrediti očekivanu vrijednost i varijansu posmatranog slučajnog signala.

- Pretpostavimo da je $P\{x(n) = 5\} = A$ za $k = 5$. Tada su vjerovatnoće da $x(n)$ uzima vrijednosti k predstavljene u donjoj tabeli:

$\xi_k = k$	0	1	2	3	4	5
$P_{x(n)}(\xi_k) = P\{x(n) = k\}$	$32A$	$16A$	$8A$	$4A$	$2A$	A

- Konstanta A se može odrediti iz

$$\sum_{k=0}^{5} P_{x(n)}(\xi_k) = 1.$$

Njena vrijednost je $A = 1/63$. Sada imamo:

$$\mu_x(n) = \mu_x = \sum_{k=0}^{5} k P_{x(n)}(\xi_k) = \frac{19}{21} \quad \text{i} \quad \sigma_x^2(n) = \sigma_x^2 = \sum_{k=0}^{5} \left(k - \frac{19}{21}\right)^2 P_{x(n)}(\xi_k) = \frac{626}{441}.$$

Funkcija gustine vjerovatnoće

- Ako slučajni signal može uzeti **kontinualne vrijednosti** po amplitudi, tada se ne može definisati vjerovatnoća da odbirak signala $x(n)$ može uzeti jednu egzaktnu vrijednost amplitude ξ.
- U ovom slučaju, koristi se **funkcija gustine vjerovatnoće** $p_{x(n)}(\xi)$ (engl. *probability density function*).
- Ova funkcija definiše vjerovatnoću da n-ti odbirak signala $x(n)$ uzima vrijednost unutar infinitezimalno malog intervala $d\xi$ oko ξ,

$$\mathrm{P}\{\xi \leq x(n) < \xi + d\xi\} = p_{x(n)}(\xi)d\xi.$$

- Funkcija gustine vjerovatnoće $p_{x(n)}(\xi)$ posjeduje sljedeća svojstva:
  1. Nenegativnost, $p_{x(n)}(\xi) \geq 0$ za bilo koje ξ
  2. Pošto je $\mathrm{P}\{-\infty < x(n) < \infty\} = 1$, to mora važiti:

$$\int_{-\infty}^{\infty} p_{x(n)}(\xi)d\xi = 1.$$

- Vjerovatnoća da je vrijednost signala $x(n)$ unutar intervala $a < x(n) \leq b$ je data izrazom:

$$\mathrm{P}\{a < x(n) \leq b\} = \int_{a}^{b} p_{x(n)}(\xi)d\xi.$$

Funkcija gustine vjerovatnoće

- **Kumulativna funkcija raspodjele** (engl. *cumulative probability distrubution – CDF*) $F_x(\chi)$ se definiše kao vjerovatnoća da je vrijednost signala $x(n)$ manja od χ,

$$F_x(\chi) = \mathrm{P}\{x(n) \le \chi\} = \int_{-\infty}^{\chi} p_{x(n)}(\xi)d\xi.$$

- Očigledno, važi $\lim_{\chi\to-\infty} F_x(\chi) = 0$, $\lim_{\chi\to+\infty} F_x(\chi) = 1$, kao i:

$$\mathrm{P}\{a < x(n) \le b\} = \int_{a}^{b} p_{x(n)}(\xi)d\xi = F_x(b) - F_x(a),$$

  gdje je $F_x(b) \ge F_x(a)$ za $b > a$.
- Kumulativna funkcija raspodjele je **neopadajuća funkcija**.
- Funkcija gustine vjerovatnoće $p_{x(n)}(\xi)$ je jednaka izvodu funkcije raspodjele $F_x(\xi)$:

$$p_{x(n)}(\xi) = \frac{dF_x(\xi)}{d\xi}.$$

- **Očekivana vrijednost** slučajne varijable $x(n)$, čiji odbirci uzimaju kontinualne vrijednosti amplitude, može se izraziti preko funkcije gustine vjerovatnoće:

$$\mu_x(n) = \mathrm{E}\{x(n)\} = \int_{-\infty}^{\infty} \xi p_{x(n)}(\xi)d\xi.$$

Funkcija gustine vjerovatnoće

- Za slučajne signale čiji odbirci uzimaju kontinualne vrijednosti amplitude **varijansa** se definiše izrazom:

$$\sigma_x^2(n) = \int_{-\infty}^{\infty} |\xi - \mu_x(n)|^2 \, p_{x(n)}(\xi) d\xi,$$

gdje je $p_{x(n)}(\xi)$ funkcija gustine raspodjele.

Primjer 6

Razmatra se realni slučajni signal $x(n)$, čiji su odbirci uniformno distribuirani na intervalu $-1 \leq x(n) < 1$.

1. Odrediti očekivanu vrijednost i varijansu odbiraka signala $x(n)$.
2. Signal $y(n)$ se dobija kao $y(n) = x^2(n)$. Odrediti očekivanu vrijednost i varijansu signala $y(n)$.

- Budući da je slučajni signal $x(n)$ uniformno distribuiran na intervalu $-1 \leq x(n) < 1$, njegova funkcija gustine vjerovatnoće je data izrazom:

$$p_{x(n)}(\xi) = \begin{cases} A & \text{za } -1 \leq \xi < 1 \\ 0 & \text{drugdje.} \end{cases}$$

Primjer 6 – rješenje

- Vrijednost konstante A, $A = 1/2$, se može jednostavno dobiti iz $\int_{-\infty}^{\infty} p_{x(n)}(\xi)d\xi = 1$.
- Očekivana vrijednost i varijansa su date izrazima:

$$\mu_x(n) = \int_{-\infty}^{\infty} \xi p_{x(n)}(\xi)d\xi = \int_{-1}^{1} \frac{1}{2}\xi d\xi = 0,$$

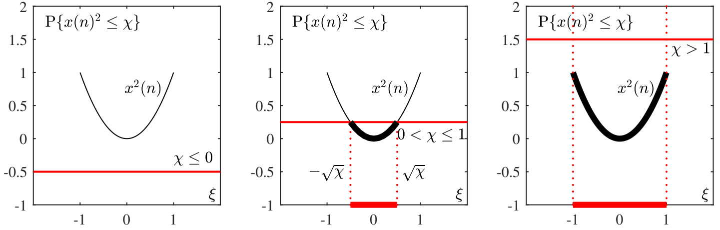
$$\sigma_x^2(n) = \int_{-\infty}^{\infty} (\xi - \mu_x(n))^2 p_{x(n)}(\xi)d\xi = \int_{-1}^{1} \frac{1}{2}\xi^2 d\xi = \frac{1}{3}.$$

- Vjerovatnoća da amplituda signala $y(n)$ nije veća od pretpostavljene vrijednosti χ predstavlja, po definiciji, distribuciju vjerovatnoće za $y(n)$. Ona je data sljedećim izrazom:

$$F_y(\chi) = P\{y(n) \leq \chi\} = P\{x^2(n) \leq \chi\} = P\{-\sqrt{\chi} \leq x(n) \leq \sqrt{\chi}\}$$
$$= \begin{cases} 0 & \text{za } \chi \leq 0 \\ \int_{-\sqrt{\chi}}^{\sqrt{\chi}} p_{x(n)}(\xi)d\xi & \text{za } 0 < \chi \leq 1 \\ 1 & \text{za } \chi > 1 \end{cases} = \begin{cases} 0 & \text{za } \chi \leq 0 \\ \sqrt{\chi} & \text{za } 0 < \chi \leq 1 \\ 1 & \text{za } \chi > 1, \end{cases}$$

pošto je $y(n) \leq \chi$, kada $x^2(n) \leq \chi$, kao što je ilustrovano na slici (naredni slajd).

Primjer 6 – rješenje



Slika: Ilustracija računanja distribucije vjerovatnoće $F_y(\chi)$ za $y(n) = x^2(n)$ i $-1 \le x(n) \le 1$.

- Funkcija gustine vjerovatnoće se dobija kao izvod od funkcije raspodjele $F_y(\chi)$, odnosno:

$$p_{y(n)}(\xi) = \frac{dF(\xi)}{d\xi} = \begin{cases} \frac{1}{2\sqrt{\xi}} & \text{za } 0 < \xi \le 1 \\ 0 & \text{inače.} \end{cases}$$

- Očekivana vrijednost i varijansa signala $y(n)$ je:

$$\mu_y(n) = \mu_y = \int_0^1 \xi \frac{1}{2\sqrt{\xi}} d\xi = \frac{1}{3}, \quad \sigma_y^2(n) = \sigma_y^2 = \int_0^1 (\xi - \frac{1}{3})^2 \frac{1}{2\sqrt{\xi}} d\xi = \frac{4}{45}.$$

Funkcija gustine vjerovatnoće – primjer

Primjer 7

- Odrediti funkciju gustine vjerovatnoće za $y(n)$ u slučaju proizvoljne monotone funkcije $y(n) = f(x(n))$ čija je inverzna funkcija $f^{-1}(y(n)) = x(n)$, pri čemu je funkcija gustine vjerovatnoće od $x(n)$ jednaka $p_{x(n)}(\xi)$.
- Odrediti izraz za funkciju gustine vjerovatnoće $p_{y(n)}(\xi)$, ako je $x(n)$ slučajna varijabla sa uniformnom funkcijom raspodjele unutar intervala $[-\pi/2, \pi/2)$ dok je $y(n) = \tan(x(n))$, dok je odgovarajuća inverzna funkcija $x(n) = \arctan(y(n))$?

- Distribucija vjerovatnoće slučajnog signala $y(n)$ data je sljedećim izrazom:

$$F_y(\chi) = P\{y(n) \le \chi\} = P\{f(x(n)) \le \chi\} = P\{x(n) \le f^{-1}(\chi)\} = \int_{-\infty}^{f^{-1}(\chi)} p_{x(n)}(\xi) d\xi,$$

odnosno

$$F_y(\chi) = F_x(f^{-1}(\chi)).$$

- Funkcija gustine vjerovatnoće je:

$$p_{y(n)}(\xi) = \frac{dF_y(\xi)}{d\xi} = \frac{dF_x(f^{-1}(\xi))}{d\xi} = p_{x(n)}(f^{-1}(\xi)) \left| \frac{df^{-1}(\xi)}{d\xi} \right|.$$

Primjer 7 – rješenje

- Ova relacija se može takođe dobiti na osnovu činjenice da vjerovatnoća u diferencijalno maloj površini mora biti invarijantna na smjenu varijabli, tj.

$$|p_{y(n)}(\xi)d\xi| = |p_{x(n)}(f^{-1}(\xi))df^{-1}(\xi)|.$$

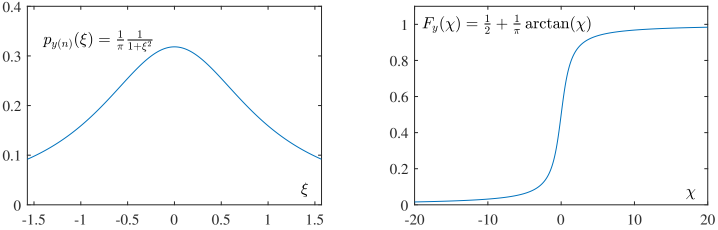
- Za slučajni signal $x(n)$, sa uniformnom funkcijom gustine vjerovatnoće unutar intervala $[-\pi/2, \pi/2)$, slučajna varijabla $y(n) = \tan(x(n))$ je distribuirana od $-\infty$ do ∞. Njena funkcija gustine vjerovatnoće je:

$$p_{y(n)}(\xi) = p_{x(n)}(\arctan(\xi))\left|\frac{d(\arctan(\xi))}{d\xi}\right| = \frac{1}{\pi}\frac{1}{1+\xi^2},$$

  pošto je $p_{x(n)}(\xi) = 1/\pi$, za $-\pi/2 \leq \xi < \pi/2$.

- Slučajni signal $y(n)$ može uzeti velike vrijednosti sa značajnom vjerovatnoćom, pošto je funkcija raspodjele: $F_y(\chi) = (1/2 + \arctan(\chi)/\pi)$.
- Slučajni signali čija raspodjela vjerovatnoće (CDF) nije eksponencijalno ograničena, odnosno, koji imaju teže (veće) vrijednosti tzv."repova" (engl. *heavy tails*), poznati su kao signali sa **heavy-tailed funkcijom gustine vjerovatnoće**.
- U ovom slučaju, zadovoljen je uslov za *heavy-tailed* funkciju $\lim_{\chi\to\infty}(F_y(\infty) - F_y(\chi))e^{a\chi} = \infty$ za svako $a > 0$.

Heavy-tailed distribucije



Slika: Ilustracija *heavy-tailed* funkcije gustine vjerovatnoće $p_{y(n)}(\xi)$ i kumulativne funkcije raspodjele $F_y(\chi)$, gdje je $y(n) = \tan(x(n))$ i $x(n) = \arctan(y(n))$.

- Po definiciji, *heavy-tailed* distribucija ima ,,rep" koji je veći od eksponencijalne funkcije, odnosno, ovakva distribucija teži nuli sporije od one koja je eksponencijalna.
- *Heavy-tailed* distribucije tipično imaju više *outlier*-a sa velikim vrijednostima, što je proporcionalno ,,težini" repa, i za njih ne postoje neki momenti.
- Neki predstavnici *heavy-tailed* distribucija su: Košijeva, Studentova, LogNormal distribucija, i druge.
- Nasuprot njima, postoje i *light-tailed* distribucije, koje imaju ,,rep" koji je,,lakši", odnosno, niži, od eksponencijalne. Primjeri su normalna i *t*-distribucija.

Združena funkcija gustine vjerovatnoće

- Kao uvod u statistiku drugog reda, posmatrajmo dva signala, $x(n)$ i $y(n)$, sa kontinualnim vrijednostima amplituda.
- Vjerovatnoća da n-ti odbirak signala $x(n)$ uzima vrijednost unutar intervala $\xi \le x(n) < \xi + d\xi$ i da $y(m)$ uzima vrijednost unutar $\zeta \le y(m) < \zeta + d\zeta$ je:

$$\mathrm{P}\{\xi \le x(n) < \xi + d\xi,\ \ \zeta \le y(m) < \zeta + d\zeta\} = p_{x(n),y(m)}(\xi,\zeta)d\xi d\zeta,$$

  gdje je $p_{x(n),y(m)}(\xi,\zeta)$ **združena funkcija gustine vjerovatnoće**.
- Vjerovatnoća događaja $a < x(n) \le b$ i $c < y(m) \le d$ je:

$$\mathrm{P}\{a < x(n) \le b,\ \ c < y(m) \le d\} = \int_a^b \int_c^d p_{x(n),y(m)}(\xi,\zeta)d\xi d\zeta.$$

- Za **međusobno nezavisne** signale važi: $p_{x(n),y(m)}(\xi,\zeta) = p_{x(n)}(\xi)p_{y(m)}(\zeta)$. Specijalni slučaj prethodnih relacija se dobija za $y(m) = x(m)$.

Primjer 8

Signal $x(n)$ je definisan kao $x(n) = a(n) + b(n) + c(n)$, gdje su $a(n)$, $b(n)$, i $c(n)$ međusobno nezavisni slučajni signali sa uniformnom funkcijom gustine vjerovatnoće na intervalu $[-1,1)$. Odrediti funkciju gustine vjerovatnoće signala $x(n)$, njegovu srednju vrijednost μ_x i varijansu σ_x^2.

Primjer 8 – rješenje

- Posmatrajmo prvo sumu dva nezavisna slučajna signala $s(n) = a(n) + b(n)$.
- Vjerovatnoća da važi $s(n) = a(n) + b(n) \leq \chi$ može biti izračunata na osnovu združene funkcije raspodjele slučajnih signala $a(n)$ i $b(n)$:

$$F_s(\chi) = P\{s(n) \leq \chi\} = \mathrm{P}\{-\infty < a(n) < \infty \text{ i } -\infty < a(n) + b(n) \leq \chi\}$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\chi-\xi} p_{a(n),b(n)}(\xi,\zeta) d\xi d\zeta = \int_{-\infty}^{\infty} \left(p_{a(n)}(\xi) \int_{-\infty}^{\chi-\xi} p_{b(n)}(\zeta) d\zeta \right) d\xi.$$

- Sada se može izračunati funkcija gustine raspodjele za $s(n)$, kao izvod od $F_s(\chi)$, odnosno:

$$p_{s(n)}(\chi) = \frac{dF_s(\chi)}{d\chi} = \int_{-\infty}^{\infty} \left(p_{a(n)}(\xi) \frac{d}{d\chi} \Big(\int_{-\infty}^{\chi-\xi} p_{b(n)}(\zeta) d\zeta \Big) \right) d\xi$$
$$= \int_{-\infty}^{\infty} p_{a(n)}(\xi) p_{b(n)}(\chi - \xi) d\xi = p_{a(n)}(\chi) *_{\chi} p_{b(n)}(\chi),$$

što znači da je funkcija gustine raspodjele zbira dvije nezavisne slučajne varijable jednaka konvoluciji pojedinačnih funkcija raspodjele.

Primjer 8 – rješenje

- Na sličan način, u slučaju tri signala se dobija da se funkcija gustine raspodjele može izračunati kao konvolucija:

$$p_{x(n)}(\chi) = p_{a(n)}(\chi) *_\chi p_{b(n)}(\chi) *_\chi p_{c(n)}(\chi),$$

$$p_{x(n)}(\chi) = \begin{cases} \frac{(\chi+3)^2}{16} & \text{for } -3 \leq \chi \leq -1 \\ \frac{3-\chi^2}{8} & \text{za } -1 < \chi \leq 1 \\ \frac{(\chi-3)^2}{16} & \text{za } 1 < \chi \leq 3 \\ 0 & \text{za } |\chi| > 3. \end{cases}$$

- Srednja vrijednost i varijansa se mogu izračunati pomoću funkcije gustine raspodjele $p_{x(n)}(\chi)$, ili direktnim putem, uz korišćenje svojstva linarnosti:

$$\begin{aligned} \mu_x &= \mathrm{E}\{x(n)\} = \mathrm{E}\{a(n)\} + \mathrm{E}\{b(n)\} + \mathrm{E}\{c(n)\} = 0 \\ \sigma_x^2 &= \mathrm{E}\{(x(n)-\mu_x)^2\} = \mathrm{E}\{(a(n)+b(n)+c(n))^2\} \\ &= \mathrm{E}\{a^2(n)\} + \mathrm{E}\{b^2(n)\} + \mathrm{E}\{c^2(n)\} + 2\left(\mu_a\mu_b + \mu_a\mu_c + \mu_b\mu_c\right) \\ &= \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1. \end{aligned}$$

Združena funkcija gustine vjerovatnoće – primjer 9

Primjer 9

Razmatraju se dva nezavisna slučajna signala $x(n)$ i $y(n)$, sa funkcijama gustine vjerovatnoće $p_{x(n)}(\xi)$ i $p_{y(n)}(\xi)$. Novi slučajni signal se definiše tako da uzima manju od vrijednosti signala $x(n)$ i $y(n)$ svakom trenutku n, odnosno:

$$z(n) = \min\{x(n), y(n)\}.$$

- Odrediti raspodjelu vjerovatnoće i funkciju gustine vjerovatnoće slučajnog signala $z(n)$.
- Čemu je jednaka funkcija raspodjele signala $z(n)$ ako su:

$$p_{x(n)}(\xi) = \frac{1}{\beta_x} e^{-\xi/\beta_x} u(\xi) \quad \text{i} \quad p_{y(n)}(\xi) = \frac{1}{\beta_y} e^{-\xi/\beta_y} u(\xi)?$$

- Pošto slučajni signal $z(n)$ uzima manju od vrijednosti $x(n)$ i $y(n)$, vjerovatnoća da je $z(n) = \min\{x(n), y(n)\}$ manje ili jednako od pretpostavljene vrijednosti χ jednaka je vjerovatnoći da je najmanje jedan od slučajnih odbiraka $x(n)$ i $y(n)$ ispod pretpostavljene vrijednosti χ:

$$P\{z(n) \leq \chi\} = P\{\min\{x(n), y(n)\} \leq \chi\}$$
$$= 1 - P\{x(n) > \chi \ \text{ i } \ y(n) > \chi\} = 1 - P\{x(n) > \chi\}P\{y(n) > \chi\}.$$

Primjer 9 – rješenje

- Pošto je
$$P\{x(n) > \chi\} = 1 - F_x(\chi) \quad \text{i} \quad P\{y(n) > \chi\} = 1 - F_y(\chi),$$
dobijamo funkciju raspodjele slučajne varijable $z(n)$ u obliku:
$$F_z(\chi) = P\{z(n) \leq \chi\} = 1 - (1 - F_x(\chi))(1 - F_y(\chi))$$
$$= F_x(\chi) + F_y(\chi) - F_x(\chi)F_y(\chi).$$

- Funkcija gustine raspodjele se dobija kao izvod funkcije raspodjele vjerovatnoće:
$$p_{z(n)}(\xi) = \frac{dF_z(\xi)}{d\xi} = p_{x(n)}(\xi) + p_{y(n)}(\xi) - p_{x(n)}(\xi)F_y(\xi) - F_x(\xi)p_{y(n)}(\xi)$$
$$= p_{x(n)}(\xi)(1 - F_y(\xi)) + p_{y(n)}(1 - F_x(\xi)).$$

- Za specifične funkcije gustine raspodjele iz postavke zadatka, dobija se:
$$p_{z(n)}(\xi) = \frac{1}{\beta_z} e^{-\xi/\beta_z} u(\xi)$$
sa $\frac{1}{\beta_z} = \frac{1}{\beta_x} + \frac{1}{\beta_y}$, pošto je:
$$F_x(\xi) = (1 - \exp(-\xi/\beta_x))u(\xi) \quad \text{i} \quad F_y(\xi) = (1 - \exp(-\xi/\beta_y))u(\xi)$$

Dodatak

Izvod integrala sa varijabilnom granicom

Napomenimo da po drugoj fundamentalnoj teoremi matematičke analize, po kojoj je $\frac{d}{dx}\int_a^x h(t)dt = h(x)$, i pravilu ulančavanja, važi sljedeće:

$$\frac{d}{dx}\int_a^{f(x)} h(t)dt = h(f(x))f'(x),$$

gdje je pretpostavljeno da su odgovarajuće funkcije neprekidne i diferencijabilne, dok a predstavlja konstantu.