

**Shannon-ове теорије неодређености и
информације
★ скрипта ★**

Виктор А. Обуљен

2006. година, Београд

Поглавље 1

Неодређеност и информација

1.1 Искуствено заснивање неодређености

Често смо у положају да погађамо један број из неког скупа бројева. Претпоставимо да у скупу има N бројева и да нам је дозвољено да постављамо питања која дозвољавају само одговоре ”да” и ”не”. Природно се намеће размишљање о томе колико нам је најмање потребно питања да бисмо погодили случајно одабрани број из задатог скупа. Једна од познатих стратегија је да се скуп подели на два (једнака, ако је могуће) дела. Сваки од новодобијених делова делимо, поново, на исти начин, на два дела и тако док у скуповима који су плод последње поделе има тачно један број.

Сваки пут стављамо 1 за ”леви” и 0 за ”десни” подскуп од добијена два. На крају сваки од бројева, тј. сваки од крајњих подскупова, има своју ”слику” изражену нулама и јединицама.

Питања постављамо, сада, пратећи ове поделе скупа свих бројева. Одговор ”да” означавамо са 1 а ”не” са 0. Решење је број који је придружен бинарном броју који смо добили.

На примеру са 5-очланим скупом лако видимо све о чему смо говорили:

$$\begin{aligned}\{1, 2, 3, 4, 5\} &\mapsto \{\{1, 2\}[1], \{3, 4, 5\}[0]\} \\ &\mapsto \{\{1\}[11], \{2\}[10], \{3\}[01], \{4, 4\}[00]\} \\ &\mapsto \{\{1\}[11], \{2\}[10], \{3\}[01], \{4\}[001], \{5\}[000]\}.\end{aligned}$$

Одговарајућа питања која бисмо постављали су:

- Да ли је број мањи од 3?
- Да ли је број мањи од 2? и Да ли је број мањи од 4?
- Да ли је број мањи од 5?

Највећа дужина бинарног броја је управо минимални број питања потребних за сигурно погађање случајног броја.

Задатак који смо себи поставили на почетку, сада, постаје проналажење највећег потребног броја бинарних цифара за једнозначно придруживање. Лако је видети да је то за један увећан цео део бинарног логаритма броја N тј. $\lceil \log_2 N \rceil + 1$.

Са становишта теорије вероватноће имамо опит са извлачењем једне од N цедуља из шешира. Пошто су нам све цедуље једнако "важне" њихове вероватноће су, све, $p = 1/N$. Број потребних питања, сада, постаје

$$-\log_2 p = -\log_2 \frac{1}{N} = \sum_{k=1}^N -\frac{1}{N} \log_2 \frac{1}{N}.$$

Задатак можемо уопштити отклањајући услов равномерности вероватноћа, тј. свакој од цедуља c_i придружимо вероватноћу p_i . Лако је уочити да је, сада, природно, број потребних питања постао $\sum_{k=1}^N -p_i \log_2 p_i$. Проверите то на једноставнијим примерима.

Шта ако уместо одговора "да" и "не" користимо вагу са два таса? Тада је могући број одговора, тј. могући су одговори: "леви тас је тежи", "десни тас је тежи" и "тасови су једнако тешки". Тада скупове делимо на три (што је могуће ближе, једнака) подскупа а логаритам "добија" основу 3.

Очекивани број питања потребних за погађање унапред задатог броја, на начин који ћемо касније детаљније описати, одговара мери коју ћемо звати неодређеност.

Да бисмо правилно дефинисали неодређеност, потребно је да направимо списак особина које та мера мора да задовољава, у складу са искуством које имамо у погађању, на начин како смо горе описали.

- Што је скуп могућих бројева већи, потребно је више питања да би погодили унапред одабрани број.
- Потребно је више питања ако је у питању равномерна расподела него када је у питању било која друга расподела вероватноћа над бројевима.
- Ако је потребно погодити "број", тј. уређен пар (састављен из два дела од којих један бирамо из скупа са N_1 а други из скупа са N_2 бројева), тада је потребан број питања једнак збиру броја питања за свако од појединачних погађања.

1.2 Аксиоматско заснивање неодређености

Нека је опит којем меримо неодређеност представљен случајно променљивом X са расподелом $\mathbb{P}\{X = x_i\} = p_i$, $i = 0, 1, \dots, N$. Показаћемо да је

$$\sum_{i=1}^N -p_i \log_2 p_i$$

мера неодређености која задовољава све горе постављене искуствене захтеве.

Пре него што то учинимо запишимо те захтеве строго. Испратимо како је те аксиоме задао Ash.

Прво, договоримо се да неодређеност означавамо са H у случају када зависи од вектора расподеле вероватноћа тј. са h када су вероватноће једнаке тј. када је у питању равномерна расподела вероватноћа.

1. $h(N)$ је монотono растућа функција од N , $N \in \mathbb{N}$.
2. $h(N_1 N_2) = h(N_1) + h(N_2)$, $N_1, N_2 \in \mathbb{N}$.
- 3.

$$\begin{aligned} H(p_1, \dots, p_N) &= H((p_1 + \dots + p_l), (p_{l+1} + \dots + p_N)) + \\ &+ (p_1 + \dots + p_l) H\left(\frac{p_1}{\sum_{i=1}^l p_i}, \dots, \frac{p_l}{\sum_{i=1}^l p_i}\right) \\ &+ (p_{l+1} + \dots + p_N) H\left(\frac{p_{l+1}}{\sum_{i=l+1}^N p_i}, \dots, \frac{p_N}{\sum_{i=l+1}^N p_i}\right) \\ &(l = 1, 2, \dots, N - 1). \end{aligned}$$

Ову аксиому називамо аксиомом груписања.

4. $H(p, 1-p)$ је непрекидна функција аргумента p . Ову аксиому називамо аксиомом непрекидности.

Теорема 1.2.1 *Једина функција која задовољава дате аксиоме је:*

$$H(p_1, \dots, p_N) = -C \sum_{i=1}^N p_i \log_b p_i, \quad (1.1)$$

где је C прозвољан позитиван реалан број, b је реалан број већи од 1.

Доказ 1.2.1 *Изведимо доказ следећи Ash-а. Једноставно је проверити да (1.1) описује функцију која задовољава све постављене аксиоме.*

Покажимо да свака функција која задовољава аксиоме 1 до 4 мора бити облика (1.1). Поделитемо то закључивање у неколико делова.

- a) $h(N^l) = lf(N)$ за све позитивне целе бројеве M и l .

То се непосредно закључује индукцијом користећи се аксиомом 2. Ако је N произвољан али задат позитиван цео број, тада (a) важи за $l = 1$. На основу аксиоме 2 је $h(N^l) = h(N \cdot N^{l-1}) = h(N) + h(N^{l-1})$ те индуктивна претпоставка да је (a) тачно за све целе бројеве укључујући и $l - 1$ имамо $h(N^l) = h(N) + (l - 1)h(N) = lh(N)$, чиме је доказ за (a) завршен.

- b) $h(N) = C \cdot \log M$ ($M = 1, 2, \dots$) где је C позитиван број.

Прво нека је $M = 1$. На основу аксиоме 2 имамо $h(1) = h(1 \cdot 1) = h(1) + h(1)$ те је $f(1) = 0$, као што и тврди (b). (Нема неодређености у погађању броја из једночланог скупа.)

Нека је, даље, N цео број већи од 1. ако је l произвољан позитиван цео број тада број 2^l лежи између нека два степена броја N тј. постоји позитиван цео број k такав да је $N^k \leq h(2^l) < M^{l+1}$. На основу аксиоме 1 следи да је $h(M^l) \leq h(2^l) < f(m^{l+1})$ и, на основу (а), важи $kh(N) \leq lh(2) < (k+1)h(N)$ тј. $k/l \leq h(2)/f(N) < (k+1)/l$. Логаритам је монотонно растућа функција (све док је основа већа од 1) те добијемо $\log N^k \leq \log 2^l < \log N^{k+1}$, тј. $k \log N \leq r \log 2 < (k+1) \log N$, тј. $k/r \leq (\log 2)/(\log N) < (k+1)/l$. Запажајући да су $h(2)/h(N)$ и $(\log 2)/\log N$ оба између k/l и $(k+1)/l$ добијемо

$$\left| \frac{\log 2}{\log N} - \frac{h(2)}{h(N)} \right| < \frac{1}{r}.$$

На основу произвољности N и l можемо дозволити да $l \rightarrow \infty$ што нам даје $(\log 2)/(\log N) = h(2)/h(N)$ тј. $h(N) = C \log N$ где је $C = h(2)/\log 2$. C мора бити позитивно јер је $h(1) = 0$ и $h(N)$ расте са порастом N .

ц) $H(p, 1-p) = -C[p \log p + (1-p) \log(1-p)]$ ако је p рационалан број.

Нека је $p = r/s$ где су r и s позитивни цели бројеви. На основу аксиоме груписања имамо да је

$$\begin{aligned} h(s) &= H\left(\underbrace{\frac{1}{s}, \dots, \frac{1}{s}}_r, \underbrace{\frac{1}{s}, \dots, \frac{1}{s}}_{s-r}\right) \\ &= H\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s}h(r) + \frac{s-r}{s}h(s-r). \end{aligned}$$

На основу (б) имамо

$$C \log s = H(p, 1-p) + Cp \log r + C(1-p) \log(s-r).$$

Следи да је

$$\begin{aligned} H(p, 1-p) &= -C [p \log r - \log s + (1-p) \log(s-r)] \\ &= -C [p \log r - p \log s + p \log s - \log s + \\ &\quad + (1-p) \log(s-r)] \\ &= -C \left[p \log \frac{r}{s} + (1-p) \log \frac{s-r}{s} \right] \\ &= -C [p \log p + (1-p) \log(1-p)]. \end{aligned}$$

д) $H(p, 1-p) = -C[p \log p + (1-p) \log(1-p)]$, за све $p \in [0, 1]$.

Горње тврђење следи непосредно из (ц) и аксиоме непрекидности. Ако је p било који број између 0 и 1 из непрекидности следи да је

$$H(p, 1-p) = \lim_{p' \rightarrow p} H(p', 1-p').$$

Посебно, можемо дозволити да се p' приближава p низом рационалних бројева. Тада имамо

$$\begin{aligned}\lim_{p' \rightarrow p} H(p', 1 - p') &= \lim_{p' \rightarrow p} [-C(p' \log p' + (1 - p') \log(1 - p'))] \\ &= -C[p \log p + (1 - p) \log(1 - p)].\end{aligned}$$

$$e) H(p_1, \dots, p_N) = -C \sum_{i=1}^N p_i \log p_i \quad (N = 1, 2, \dots).$$

Већ смо установили да горња формула важи за $N = 1$ и $N = 2$. За $N > 2$, користећи аксиому 3 добијамо

$$\begin{aligned}H(p_1, \dots, p_N) &= H(p_1 + \dots + p_{N-1}, p_N) + \\ &+ (p_1 + \dots + p_{N-1}) \cdot H\left(\frac{p_1}{\sum_{i=1}^{N-1} p_i}, \dots, \frac{p_{N-1}}{\sum_{i=1}^{N-1} p_i}\right) + p_N H(1).\end{aligned}$$

Ако претпоставимо да је формула тачна за све позитивне целе бројеве до $N - 1$ добијамо

$$\begin{aligned}H(p_1, \dots, p_N) &= -C[(p_1 + \dots + p_{N-1}) \log(p_1 + \dots + p_{N-1}) \\ &+ p_N \log p_N] - C(p_1 + \dots + p_{N-1}) \\ &\quad \left[\frac{p_1}{\sum_{i=1}^{N-1} p_i} \log\left(\frac{p_1}{\sum_{i=1}^{N-1} p_i}\right) + \dots + \right. \\ &\quad \left. + \frac{p_{N-1}}{\sum_{i=1}^{N-1} p_i} \log\left(\frac{p_{N-1}}{\sum_{i=1}^{N-1} p_i}\right) \right] + p_N(0) \\ &= -C \left[\left(\sum_{i=1}^{N-1} p_i \right) \log \left(\sum_{i=1}^{N-1} p_i \right) + p_N \log p_N \right] \\ &= -C \left[\sum_{i=1}^{N-1} p_i \log p_i - \left(\sum_{i=1}^{N-1} p_i \right) \log \sum_{i=1}^{N-1} p_i \right] \\ &= -C \sum_{i=1}^N p_i \log p_i.\end{aligned}$$

Тиме је доказ завршен. □

1.3 Особине неодређености

У наредним одељцима ћемо пратити Yeung-а.

Скоро најважнија особина неодређености јесте да се, у случају када неодређеност рачунамо за случајно променљиву X неодређеност представља

$$H(X) := - \sum_x p(x) \log p(x) = -\mathbb{E}(\log p(X)) \quad (1.2)$$

где смо са $p(X)$ означили закон расподеле случајно променљиве X . Неодређеност можемо посматрати и за случајне векторе, тј. парове случајно променљивих. У том случају имамо

$$H(X, Y) := - \sum_{x, y} p(x, y) \log p(x, y) = -\mathbb{E} \log p(X, Y). \quad (1.3)$$

Настављајући даље можемо посматрати и

$$H(Y|X) := - \sum_{x, y} p(x, y) \log p(y|x) = -\mathbb{E} \log p(Y|X). \quad (1.4)$$

тј.

$$H(Y|X) = \sum_x p(x) \left[- \sum_y p(y|x) \log p(y|x) \right], \quad (1.5)$$

где је

$$- \sum_y p(y|x) \log p(y|x) =: H(Y|X = x) \quad (1.6)$$

па имамо

$$H(Y|X) = \sum_x p(x) H(Y|X = x). \quad (1.7)$$

Став 1.3.1

$$H(X, Y) = H(X) + H(Y|X) \quad (1.8)$$

$$H(X, Y) = H(Y) + H(X|Y). \quad (1.9)$$

Доказ 1.3.1 Користићемо скраћено обележавање које смо увели раније.

$$\begin{aligned} H(X, Y) &= -\mathbb{E} \log p(X, Y) \\ &= -\mathbb{E} \log [p(X)p(Y|x)] \\ &= -\mathbb{E} \log p(X) - \mathbb{E} \log p(X|Y) \\ &= H(X) + H(Y|X). \end{aligned}$$

□

Тумачење претходног става је да се погађање исхода пара случајно променљивих (X, Y) одвија у два корака: прво погађамо исход случајно променљиве X а затим исход случајно променљиве Y . Теорема каже да је укупна количина неодређености једнака неодређености уклоњеној приликом погађања исхода X увећаној за неодређеност за Y при услову да знамо исход X .

Претходни став се може уопштити на следећи начин:

$$\begin{aligned} H(X, Y, Z) &= H(X) + H(Y|X) + H(Z|X, Y) \\ &= H(X, Y) + H(Z|X, Y) \\ &= H(X) + H(Y, Z|X), \end{aligned}$$

т.ј.

$$H(X_1, \dots, X_n, Y_1, \dots, Y_m) = H(X_1, \dots, X_n) + H(Y_1, \dots, Y_m | X_1, \dots, X_n).$$

Једна, иако веома једноставна, од, слободно можемо рећи, најважнијих неједнакости у теорији информације дата је следећим ставом.

Став 1.3.2 (Shannon-ова неједнакост) *Нека су $(p_i)_{i=1}^N$ и $(q_i)_{i=1}^N$ произвољни позитивни бројеви који задовољавају $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i = 1$. Тада је*

$$-\sum_{i=1}^N p_i \log p_i \leq -\sum_{i=1}^N p_i \log q_i$$

уз једнакост ако и само ако је $p_i = q_i$ за све i .

Доказ 1.3.2 *Доказ почињемо везом: $\log_2 x = \log_2 e \log_e x = \log_2 e \cdot \ln x$, која обезбеђује независност горње неједнакости од основе логаритма.*

С друге стране особина \ln нам поједностављује доказ. Наиме, $\ln x \leq x - 1$ уз једнакост ако и само ако $x = 1$. Дакле, $\ln(q_i/p_i) \leq q_i/p_i - 1$ уз једнакост ако и само ако $p_i = q_i$. Множећи са p_i и сабирајући по i добијамо

$$\sum_{i=1}^N p_i \ln \frac{q_i}{p_i} \leq \sum_{i=1}^N (q_i - p_i) = 1 - 1 = 0$$

уз једнакост ако и само ако $p_i = q_i$ за све i . Дакле,

$$\sum_{i=1}^N p_i \ln q_i - \sum_{i=1}^N p_i \ln p_i \leq 0.$$

□

Користећи Shannon-ову неједнакост можемо доказати, на пример, и неједнакост аритметичке и геометријске средине

$$x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \leq \sum_{i=1}^n a_i \cdot x_i.$$

Покушајте.

Ево и доказа особине који смо дуговали од почетних заснивања неодређености.

Став 1.3.3

$$H(p_1, \dots, p_N) \leq \log N = h(N) = H(1/N, \dots, 1/N)$$

уз једнакост ако и само ако је

$$p_i = 1/N, \quad i = 1, \dots, N.$$

□

Доказ 1.3.3 *Применом претходног става са $q_i = 1/N$ добијамо*

$$-\sum_{i=1}^N p_i \log p_i \leq -\sum_{i=1}^N p_i \log \frac{1}{N} = \log N \sum_{i=1}^N p_i = \log N$$

уз једнакост ако и само ако је

$$p_i = q_i = 1/N, \quad i = 1, \dots, N.$$

Став 1.3.4 $H(X, Y) \leq H(X) + H(Y)$ *уз једнакост ако и само ако су X и Y независне.*

Доказ 1.3.4 *Запажајући да је*

$$p(x_i) = \sum_{j=1}^{N_2} p(x_i, y_j) \quad \text{и} \quad p(y_j) = \sum_{i=1}^{N_1} p(x_i, y_j)$$

можемо записати:

$$H(X) = -\sum_{i=1}^{N_1} p(x_i) \log p(x_i) = -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p(x_i, y_j) \log p(x_i)$$

$$H(Y) = -\sum_{j=1}^{N_2} p(y_j) \log p(y_j) = -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p(x_i, y_j) \log p(y_j)$$

Одатле имамо:

$$\begin{aligned} H(X) + H(Y) &= -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p(x_i, y_j) [\log p(x_i) + \log p(y_j)] \\ &= -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p(x_i, y_j) [\log p(x_i) \cdot p(y_j)] \\ &= -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{ij} \log p(x_i) p(y_j) \end{aligned}$$

Ставимо да је $q_{ij} = p(x_i)p(y_j)$. Тада, имамо

$$H(X, Y) = -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{ij} \log p_{ij}.$$

Примењујући Шанпон-ову неједнакост добијамо

$$\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{ij} \log p_{ij} \leq -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{ij} \log q_{ij}$$

уз једнакост ако и само ако је $p_{ij} = q_{ij}$ за све i, j .

□

Слично можемо доказати да је

$$H(X_1, \dots, X_n) \leq H(X_1) + \dots + H(X_n)$$

уз једнакост ако и само ако су X_1, \dots, X_n независне.

Став 1.3.5 $H(Y|X) \leq H(Y)$ уз једнакост ако и само ако су X и Y независне.

Доказ 1.3.5 Из претходног имамо да је

$$H(X) + H(Y|X) = H(X, Y) \leq H(X) + H(Y)$$

уз једнакост ако и само ако су X и Y независне. □

Да би сте се уверили да сте савладали претходно, докажите још једну од значајних неједнакости теорије информације, да је

$$H(Y, Z|X) \leq H(Y|X) + H(Z|X).$$

1.4 Неодређеност у ланцима

Став 1.4.1

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i|X_1, \dots, X_{i-1}). \quad (1.10)$$

Доказ 1.4.1 За $n = 2$ смо тврђење већ доказали. Доказујемо даље индукцијом по n . Претпоставимо да је тврђење тачно за $n = m$, где је $m \geq 2$. Тада је

$$\begin{aligned} & H(X_1, \dots, X_m, X_{m+1}) \\ &= H(X_1, \dots, X_m) + H(X_{m+1}|X_1, \dots, X_m) \\ &= \sum_{i=1}^m H(X_i|X_1, \dots, X_{i-1}) + H(X_{m+1}|X_1, \dots, X_m) \\ &= \sum_{i=1}^{m+1} H(X_i|X_1, \dots, X_{i-1}). \end{aligned}$$

□

Став 1.4.2

$$H(X_1, X_2, \dots, X_n|Y) = \sum_{i=1}^n H(X_i|X_1, \dots, X_{i-1}, Y). \quad (1.11)$$

Доказ 1.4.2

$$\begin{aligned} & H(X_1, X_2, \dots, X_n|Y) \\ &= \sum_y p(y) H(X_1, X_2, \dots, X_n|Y = y) \\ &= \sum_{i=1}^n \sum_y H(X_1, X_2, \dots, X_n|Y = y) \\ &= \sum_{i=1}^n H(X_1, X_2, \dots, X_n|Y), \end{aligned}$$

□

1.5 Информација

Под појмом информација између две случајно променљиве подразумевамо неодређеност о вредности случајно променљиве, рецимо, X која буде отклоњена након што сазнамо вредност неке друге случајно променљиве, рецимо, Y . Једноставно $I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.

Мера информације (даље краће: информација) између случајно променљивих X и Y се задаје као:

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = \mathbb{E} \log \frac{p(X, Y)}{p(X)p(Y)}. \quad (1.12)$$

Став 1.5.1 Информација између X и X једнака је неодређености за X . тј. $I(X; X) = H(X)$.

Доказ 1.5.1

$$\begin{aligned} I(X; X) &= \mathbb{E} \log \frac{p(X)}{p(X)^2} \\ &= -\mathbb{E} \log p(X) \\ &= H(X). \end{aligned}$$

□

На тај начин скоро да све особине можемо доказивати за само једну од ове две мере. Отуд је могући назив за неодређеност и **самоинформација**.

Лако је уочити (проверите) да је информација симетрична по X и Y тј.

$$I(X; Y) = I(Y; X).$$

Отуда и не користимо у ознаци знак $|$ већ $;$.

И информација се проширује на условне расподеле вероватноће и то као

$$I(X; Y|Z) := \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = \mathbb{E} \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$

Докажите да је, и у овом случају, $I(X; Y|Z) = I(Y; X|Z)$.

Скраћено записивање је и овде веома корисно

$$I(X; Y|Z) = \sum_z p(z) I(X; Y|Z = z)$$

где је

$$I(X; Y|Z = z) := \sum_{x,y} p(x, y|z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)}.$$

Слично када је услов по две случајно променљиве пишемо

$$I(X; Y|Z, T) := \sum_t p(t) I(X; Y|Z, T = t)$$

где је

$$I(X; Y|Z, T = t) = \sum_{x, y, z} p(x, y, z|t) \frac{p(x, y|z, t)}{p(x|z, t)p(y|z, t)}.$$

Да би сте се уверили да сте савладали претходно докажете следећа два тврђења.

Став 1.5.2

$$I(X; X|Z) = H(X|Z).$$

Став 1.5.3

$$\begin{aligned} I(X; Y|Z) &= H(X|Z) - H(X|Y, Z), \\ I(X; Y|Z) &= H(Y|Z) - H(Y|X, Z), \\ I(X; Y|Z) &= H(X|Z) + H(Y|Z) - H(X; Y|Z). \end{aligned}$$

Што се тиче терминологије за неодређеност и информацију обавезно погледајте:

<http://www.lecb.ncifcrf.gov/toms/information.is.not.uncertainty.html>.

Ми смо, управо, и усвојили израз ”неодређеност” док појам ”информација” користимо као назив мере а израз ”податак” ћемо користити за оно што се, најчешће, у нашем језику, погрешно, користи појам ”информација”.

1.6 Информација у ланцима

Став 1.6.1 (Data Processing Inequality) Нека U, X, Y, V образују ланац Маркова (ознака коју ћемо, даље, користити је $U \propto X \propto Y \propto V$). Тада је

$$I(U; V) \leq I(X; Y).$$

Доказ 1.6.4 Пажљивим праћењем дефиниција и особина ланца Маркова. □

1.7 Нека друга тумачења информације

Претпоставимо да је за случајно променљиве X и Y могуће наћи скупове \tilde{X} и \tilde{Y} такве да су:

- (i) \tilde{X} и \tilde{Y} су дисјунктни ако су X и Y независне,
- (ii) $\tilde{X} = \tilde{Y}$ кадгод су X и Y једнозначно одређене једна другом, тј. у питању је обозначно једнозначно пресликавање.

Прво, проверите да ли такви скупови (увек?) постоје. Ако постоје онда је могућ успоставити следеће везе, где је μ нека (мора ли бити?) вероватносна мера:

$$\begin{aligned}
H(X) &= \mu(\tilde{X}), \\
H(Y) &= \mu(\tilde{Y}), \\
H(X, Y) &= \mu(\tilde{X} \cup \tilde{Y}), \\
H(X|Y) &= \mu(\tilde{X} \setminus \tilde{Y}), \\
I(X; Y) &= \mu(\tilde{X} \cap \tilde{Y}), \\
H(X, Y, Z, \dots) &= \mu(\tilde{X} \cup \tilde{Y} \cup \tilde{Z} \cup \dots), \\
I(X; Y; Z; \dots) &= \mu(\tilde{X} \cap \tilde{Y} \cap \tilde{Z} \cap \dots).
\end{aligned}$$

Горње везе нису се, до сада, показале значајним али су погодне за успостављање интуиције ако сте довољно добро научили теорију мере или теорију вероватноће.

1.8 Kullback-Leibler-ово растојање

Нека су p и q две вероватносне мере над заједничким вероватносним простором скупа елементарних исхода \mathfrak{X} . Често нам је потребна мера колико је p различита од q или обратно. Та мера мора задовољавати основне захтеве тј да је нула када и само када је $p = q$ и да је ненегативна. Означимо са \mathbb{E}_p очекивање у односу на p . Kullback-Leibler-ово разилажење (information divergence) дефинисано изразом

$$\mathbb{KL}(p||q) := \sum_x p(x) \frac{p(x)}{q(x)} = \mathbb{E}_p \log \frac{p(x)}{q(x)}$$

задовољава задате услове.

Став 1.8.1

$$\mathbb{KL}(p||q) \neq \mathbb{KL}(q||p).$$

Став 1.8.2

$$\mathbb{KL}(p||q) \geq 0.$$

Доказ 1.8.2 Непосредном применом Shannon-ове неједнакости. \square

Изворно Kullback и Leibler су дефинисали симетрично и ненегативно растојање: $\mathbb{KL}(P||Q) + \mathbb{KL}(Q||P)$ али оно није нашло скоро никакве примене.

Пинскер је предложио $\mathbf{d}(p, q) = \sum_{x \in \mathfrak{X}} |p(x) - q(x)|$ као меру разилажења две вероватносне мере.

За вежбу докажете следећи став и једнакости (P_u је равномерна расподела на одговарајућем скупу вредности):

Став 1.8.3

$$\begin{aligned}
\mathbb{KL}(P||Q) &= - \sum_x p(x) \log q(x) + \sum_x p(x) \log p(x) \\
&= H(P, Q) - H(P)
\end{aligned}$$

$$\mathbb{KL}(p||q) \geq \frac{1}{2 \ln 2} \mathbf{d}^2(p, q).$$

$$I(m) = \mathbb{KL}(\delta_{im} || \{p_i\}), \text{ где је } \delta_{ij} = \begin{cases} 1, & \text{за } i = j \\ 0, & \text{за } i \neq j \end{cases}.$$

$$\begin{aligned} I(X; Y) &= \mathbb{KL}(P(X, Y) || P(X)P(Y)) \\ &= \mathbb{KL}(P(Y|X) || P(Y)) \\ &= \mathbb{KL}(P(X|Y) || P(X)), \end{aligned}$$

$$H(X) = \log N - \mathbb{KL}(P(X) || P_u(X))$$

$$\begin{aligned} H(X|Y) &= \log N - \mathbb{KL}(P(X, Y) || P_u(X)P(Y)) \\ &= \log N - \mathbb{KL}((P(X, Y) || P(X)P(Y)) - \mathbb{KL}((P(X) || P_u(X))) \\ &= \log N - \mathbb{KL}(P(X|Y) || P_u(X)) \end{aligned}$$

Kullback-Leibler-ово разлижење је нашло значајну примену у математичкој статистици.

1.9 Неједнакост Fano-a

Разлоге за извођење ове особине неодређености ћемо видети када будемо доказивали обрат основних теорема кодирања теорије информације.

Став 1.9.1 *За сваку случајно променљиву X је*

$$H(X) \leq \log \#\mathbb{X},$$

где је $\#\mathbb{X}$ величина скупа могућих вредности случајно променљиве X тј. \mathbb{X} . Ова граница је тачна ако и само ако је X равномерно расподељена на \mathbb{X} .

Доказ 1.9.1 *Нека је u равномерна расподела над \mathbb{X} , тј. $u(x) = \frac{1}{\#\mathbb{X}}$ за све $x \in \mathbb{X}$. Означимо са \mathcal{S}_X скуп вредности које узима случајно променљива X . Тада је*

$$\begin{aligned} \log \#\mathbb{X} - H(X) &= - \sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{\#\mathbb{X}} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\ &= - \sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\ &= \sum_{x \in \mathcal{S}_X} p(x) \frac{p(x)}{u(x)} \\ &= \mathbb{KL}(p||u) \\ &\geq 0 \end{aligned}$$

□

Нека су X и \hat{X} случајно променљиве које узимају вредности из истог скупа (азбуке) \mathfrak{X} . Обележимо $P_e = \mathbb{P}\{X \neq \hat{X}\}$.

Теорема 1.9.1 (Неједнакост Фапо-а) *Нека су X и \hat{X} случајно променљиве које узимају вредности из исте азбуке \mathfrak{X} . Тада је*

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(\#\mathfrak{X} - 1).$$

Доказ 1.9.2 *Нека је*

$$Y = \begin{cases} 0, & X = \hat{X} \\ 1, & X \neq \hat{X} \end{cases}$$

$H(Y|X, \hat{X}) = 0$ јер је Y функција X и \hat{X} (докажите). Тада је

$$\begin{aligned} H(X|\hat{X}) &= I(X; Y|\hat{X}) + H(X|\hat{X}, Y) \\ &= H(Y|\hat{X}) - H(Y|X, \hat{X}) + H(X|\hat{X}, Y) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\leq H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathfrak{X}} [\mathbb{P}\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0) + \\ &\quad + \mathbb{P}\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)]. \end{aligned}$$

X мора узети вредност \hat{x} ако је $\hat{X} = \hat{x}$ и $Y = 0$. Одатле је

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

Ако је $\hat{X} = \hat{x}$ и $Y = 1$ тада X мора узети вредност из скупа $\{x \in \mathfrak{X} : x \neq \hat{x}\}$ који садржи $\#\mathfrak{X} - 1$ чланова. Дакле, према последњој доказаној теорему имамо

$$H(X|\hat{X} = \hat{x}, Y = 1) \leq \log(\#\mathfrak{X} - 1),$$

где ова горња граница не зависи од \hat{x} . На основу тога закључујемо:

$$\begin{aligned} H(X|\hat{X}) &\leq h_b(P_e) + \left(\sum_{\hat{x} \in \mathfrak{X}} \mathbb{P}\{\hat{X} = \hat{x}, Y = 1\} \right) \log(\#\mathfrak{X} - 1) \\ &= h_b(P_e) + \mathbb{P}\{Y = 1\} \log(\#\mathfrak{X} - 1) \\ &= h_b(P_e) + P_e \log(\#\mathfrak{X} - 1), \end{aligned}$$

□

Једноставна последица неједнакости Фапо-а је следећа неједнакост коју ћемо користити чешће од неједнакости Фапо-а саме.

Став 1.9.2

$$H(X|\hat{X}) < 1 + P_e \log \#\mathfrak{X}.$$

1.10 Неодређеност стационарног извора

Дискретан извор података $\{X_i, i = 1, \dots\}$ је бесконачна збирка случајно променљивих уређених скупом природних бројева. Случајно променљиве X_i ћемо звати словима. Претпоставимо, даље, да су све $H(X_i)$ коначне. Тада за сваки коначан подскуп A скупа обележја $\{i : i = 1, \dots\}$ посматрамо

$$H(X_i, i \in A) \leq \sum_{i \in A} H(X_i) < \infty.$$

Пошто је, осим у посебним случајевима, неодређеност бесконачне збирке бесконачна, посматрамо

$$H_X = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, \dots, X_n)$$

и ако постоји то ћемо сматрати неодређеношћу дискретног извора података.

За извор података кажемо да је **стационаран** (отпоран на проток времена који поистовећујемо са порастом обележја) ако је расподела за X_1, X_2, \dots, X_n иста као и за $X_{m+1}, X_{m+2}, \dots, X_{m+l}$.

У описивању граничног понашања за $\{X_i\}$ користићемо, ако постоји

$$H'_X = \lim_{n \rightarrow \infty} H(X_n | X_1, X_2, \dots, X_{n-1}).$$

Став 1.10.1 Нека је $\{X_i\}$ стационаран извор. Тада H'_X постоји.

Доказ 1.10.1

$$\begin{aligned} H(X_n | X_1, X_2, \dots, X_{n-1}) &\leq H(X_n | X_2, X_3, \dots, X_{n-1}) \\ &= H(X_{n-1} | X_1, X_2, \dots, X_{n-2}), \end{aligned}$$

Нерастући низ ограничен нулом одоздо је конвергентан. \square

Став 1.10.2 За сваки стационаран извор података $\{X_i\}$ постоји неодређеност H_X и једнака је H'_X .

Доказ 1.10.2

$$\frac{1}{n} H(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n H(X_i | X_1, X_2, \dots, X_{i-1})$$

На основу теореме (Cesáro) о томе да за сваки низ реалних бројева такав да је $a_n \rightarrow a, n \rightarrow \infty$ важи и да $\frac{1}{n}(a_1 + \dots + a_n) \rightarrow a, n \rightarrow \infty$ лако закључујемо да је

$$H_X = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n) = H'_X.$$

\square

Доказали смо да неодређеност постоји за извор података ако је он стационаран, но, ако тај извор није и ергодичан, тада добијена вредност не мора имати и суштинско значење. На овом месту појам ергодичности превазилази наше потребе и обрађене појмове, те ћемо се на то вратити ако и када то буде прикладно.

1.11 Подсећање

На почетку овог поглавља дали смо пример са погађањем случајно одабраног броја из петочланог скупа.

Питањима која смо постављали вредност $\log_2 5$ смо смањивали за неодређеност отклоњену одговарањем на свако од постављених питања, условно претходних одговора.

Проверите.

Ако сте помислили на то да смо неодређеност $\log_2 5$ представили као збир стечених информација на основу сваког од питања, одлично, спремни сте за следеће поглавље.

Поглавље 2

Кодирање без шумова

2.1 Код

Нека нам је дата случајно променљива X са скупом могућих вредности \mathcal{X} . Нека нам је дат коначан скуп \mathcal{D} ($\#\mathcal{D} < \infty$). Код је пресликавање $\mathcal{C} : \mathcal{X} \rightarrow \mathcal{D}^*$ где је $\mathcal{D}^* = \cup_{i=1}^{\infty} \mathcal{D}^i$.

Како кодирамо извор података? Тако што кодирамо свако слово које он произведе и те (кодне) речи (ниске) надовезујемо једну на другу по реду наилазак изворних знакова. Наравно, поставља се проблем повратка на изворни низ слова тј. декодирања. У том циљу посматрамо посебну врсту кодова, тзв. једнозначно препознатљиве. Код је једнозначно препознатљив ако сваки коначан низ (у даљем, ниска) кодних знакова одговара само једној изворној поруци тј. ниски изворних знакова (у даљем, слова). Дакле, једнозначно препознатљив код је, једноставно, обострано једнозначно пресликавање из скупа слова извора података у скуп кодних речи (ниски над кодном азбуком).

Скоро сваки извор података поседује нешто што називамо сувишношћу. То је, једноставно

$$1 - \frac{H(X)}{\log N}.$$

Вратимо се на задатак са почетка књиге. Ако имамо четири броја од којих бирамо један и ако нису сви бројеви једнако вероватни, онда, у случајевима са погодном одабраним вероватноћама, не мора бити потребно два питања за (понављам, сигурно) откривање одабраног броја. Пробајте.

Циљ кодирања без шумова јесте да се уклони сувишност ради што ефикаснијег преноса података. Наиме, код извора података са сувишношћу преносило би се више знакова него што је то потребно. Данас се кодирање без шумова, једноставно, назива сажимањем података тј. data compression.

2.2 Kraft-ова неједнакост

Теорема 2.2.1 Нека је \mathcal{C} код са кодним азбуком од D слова. Нека су, даље, l_1, l_2, \dots, l_n дужине кодних речи. Ако је \mathcal{C} једнозначно препознатљив тада је

$$\sum_{i=1}^n D^{-l_i} \leq 1.$$

Доказ 2.2.1

$$\begin{aligned} \left(\sum_{i=1}^n D^{-l_i}\right)^N &= \left(\sum_{i_1=1}^n D^{-l_{i_1}}\right) \left(\sum_{i_2=1}^n D^{-l_{i_2}}\right) \dots \left(\sum_{i_N=1}^n D^{-l_{i_N}}\right) \\ &= \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_N=1}^n D^{-(l_{i_1} + l_{i_2} + \dots + l_{i_N})} \\ &= \sum_1^{Nl_{max}} A_i D^{-i} \end{aligned}$$

где је $l_{max} = \max_i \{l_i\}$ и A_i је број кодних речи дужине N са i знакова. Лако је закључити да је $D_i \leq A_i$ јер је број могућих ниски дужине i над азбуком од D знакова не мањи од броја кодних речи дужине i . Одатле имамо:

$$\left(\sum_{i=1}^n D^{-l_i}\right)^N = \sum_1^{Nl_{max}} A_i D^{-i} = Nl_{max}$$

па је $\left(\sum_{i=1}^n D^{-l_i}\right) \leq (Nl_{max})^{1/N}$. Када $N \rightarrow \infty$ тада $(Nl_{max})^{1/N} \rightarrow 1$. \square

2.3 Префикс и тренутан код

За код кажемо да је префикс код ако ниједна кодна реч није префикс неке друге кодне речи истог кода. Лако закључујемо да је сваки префикс код и једнозначно препознатљив, тј. нема кодне ниске која би се могла препознати (декодирати) у две различите изворне речи, и тренутно препознатљив тј. препознавање се може извршити слово по слово, није потребно чекати завршетак кодне речи да би се вршило препознавање.

Теорема 2.3.1 За позитивне бројеве l_1, l_2, \dots, l_n постоји префикс код ако и само ако је за те бројеве задовољена неједнакост Kraft-а.

Доказ 2.3.1 Пошто је сваки префикс код једнозначно препознатљив, из Kraft-ове неједнакости директно следи потребност.

Без губитка опитности можемо претпоставити да је $l_1 \leq l_2 \leq \dots \leq l_n$ и посматрати све ниске дужине l_n над азбуком обима D којих има D^{l_n} . Идеја је да се почне са пуним, балансираним D -арним дрветом, изабере кодна реч и отклоне сви њени изданци са дрвета. Пошто је $D^{l_1} \geq 1$ увек можемо наћи прву кодну реч. Претпоставимо, даље, да је нађено

i ко́дних речи. Показујемо да можемо наћи нову дужине l_{i+1} . За све $1 \leq j < i$, ко́дна реч дужине j има $D^{l_{i+1}-l_j}$ изданака реда l_{i+1} на дрвету. Сада има $D^{l_{i+1}}$ чворова на пуном дрвету дубине l_{i+1} . Поткресали смо $\sum_{j=1}^i D^{l_{i+1}-l_j}$ чворова са дрвета. Број преосталих чворова на дубини $i+1$, тада, је:

$$\begin{aligned} D^{l_{i+1}} - \sum_{j=1}^i D^{l_{i+1}-l_j} &= D^{l_{i+1}} - \sum_{j=1}^{i+1} D^{l_{i+1}-l_j} + D^{l_{i+1}-l_{i+1}} \\ &= D^{l_{i+1}} - D^{l_{i+1}} \sum_{j=1}^{i+1} D^{-l_j} + 1 \\ &\geq D^{l_{i+1}}(1-1) + 1 \geq 1 \quad (\text{Kraft}) \end{aligned}$$

те увек можемо наћи ко́дну реч дужине $i+1$ ако је Kraft-ова неједнакост задовољена. \square

2.4 Оптималност ко́да

Теорема 2.4.1 (Теорема о ко́дирању без шумова) Нека је \mathcal{C} једнозначно препознатљив ко́д азбуке са D слова за случајно променљиву X . Тада је очекивана дужина ко́дне речи у \mathcal{C} одоздо ограничена неодређеношћу те случајно променљиве, $H_D(X)$ (рачунате по логаритму основе D). Та доња граница је тачна ако и само ако је $l_i = -\log_D p_i$ за све i .

Доказ 2.4.1

$$\begin{aligned} L - H_D(X) &= \sum p_i \log_D D^{l_i} + \sum p_i \log_D p_i \\ &= \sum p_i \log_D p_i D^{l_i} \\ &= \frac{1}{\ln D} \sum p_i \ln p_i D^{l_i} \\ &\geq \frac{1}{\ln D} \sum p_i \left(1 - \frac{1}{p_i D^{l_i}}\right) \quad (\ln a \leq a - 1) \\ &= \frac{1}{\ln D} \sum \left(p_i - \frac{1}{D^{l_i}}\right) \\ &= \frac{1}{\ln D} \left(1 - \sum D^{-l_i}\right) \\ &\geq \frac{1}{\ln D} (1-1) \quad (\text{Kraft}) \\ &= 0. \end{aligned}$$

Лако је проверити да за $l_i = -\log_D p_i$ важи $L = \sum p_i \log_D \frac{1}{p_i} = H_D(X)$. За доказ тачности границе потребно је да једнакост важи на два места где се појављује неједнакост. Добијено ограничење $a = \frac{1}{p_i D^{l_i}} = 1$ доказује потребност услова. \square

Потражимо начин да сачинимо префикс ко́д са најмањом очекиваном дужином ко́дне речи. Циљ нам је да минимизујемо $L = \sum p_i l_i$ уз

задато $\sum D^{-l_i} \leq 1$. Користећи Lagrange-ове мултипликаторе вршимо минимизацију

$$J = \sum p_i l_i + \lambda (\sum D^{-l_i})$$

што, након диференцирања по l_i , даје

$$\frac{\partial J}{\partial l_i} = p_i + \lambda (D^{-l_i} \log_C D).$$

Изједначавајући то са 0 добијамо

$$D^{-l_i} = \frac{p_i}{\lambda \log_C D}.$$

Користећи се ограничењем налазимо да је $D^{-l_i} = p_i$ најбоље решење, дакле $l_i = -\log_D p_i$ је најбоља дужина кодне речи. Проблем је у томе што дужина кодне речи мора бити природан број. Но, показује се да можемо постићи дужину која је блиска најбољој. Нека је $l'_i = \lceil -\log_D p_i \rceil$. Јасно је да је $\sum D^{-l'_i} \leq 1$. Тада је

$$\log_D p_i \leq l'_i < -\log_D p_i + 1 \quad \text{т.ј.}$$

$$\sum p_i \log_D p_i \leq \sum p_i l'_i < -\sum p_i \log_D p_i + p_i \quad \text{т.ј.}$$

$$H_D(X) \leq L \leq H_D(X) + 1.$$

Дакле, најбоља очекивана дужина кодне речи је највише за 1 далеко од $H_D(X)$. Остаје нам да покушамо да то и остваримо. Кренимо редом како су такви кодови настајали.

2.5 Shannon-ово бинарно кьдрање

Claude Shannon је предложио да се успостави следећа неједнакост:

$$\frac{H(X)}{\log D} \leq \bar{L} = \sum_{k=1}^N P\{x_k\} n_k < 1 + \frac{H(X)}{\log D}$$

Одатле се једноставно добија да је:

$$2^{-n_k} \leq P\{n_k\}, \quad k = \overline{1, N}.$$

Да би све ове неједначине биле испуњене мора важити:

$$\sum_{k=1}^N 2^{-n_k} \leq 1.$$

Дакле, знамо да тражени кьд постоји. Такав кьд ће имати особину да је:

$$H(X) \leq \bar{L} < H(X) + 1.$$

Алгоритам кьдирања је следећи:

1. скуп порука поставити у нерастући поредак по њиховим вероватноћама

$$[x_1, x_2, \dots, x_N] \quad : \quad P\{x_1\} \geq P\{x_2\} \geq \dots \geq P\{x_N\}$$

2.

$$\alpha_1 = 0$$

$$\alpha_2 = P\{x_1\}$$

$$\alpha_3 = P\{x_2\} + P\{x_1\} = P\{x_2\} + \alpha_2$$

$$\alpha_4 = P\{x_3\} + \alpha_3$$

...

3. наћи најмање целе $(n_i)_{i=1}^{\infty}$ као решења неједначина:

$$2^{n_i} P\{x_i\} \geq 1, \quad i = \overline{1, \infty}$$

4. развити децималан број α_i у бинарни запис до n_i -тог места.

Покажимо да је то се сасвим сагласно и да се овакво кодирање може извести. Због већ уочене чињенице да је $\sum_{k=1}^N 2^{-n_k} \leq 1$ постоји префикс кôд са жељеним дужинама кодних речи.

Даље, имамо да је $0 = \alpha_1 < \alpha_2 < \dots < \alpha_n < \alpha_{n+1} = 1$. Претпоставимо да α_k развијамо до n_k -тог места:

$$\alpha_k \rightarrow .\tau_{-1}\tau_{-2}\dots\tau_{-n_k}$$

$$\alpha_{k+1} \rightarrow .\tau_{-1}\tau_{-2}\dots\tau_{-n_{k+1}}, \quad n_{k+1} \geq n_k.$$

Но, $\alpha_{k+1} = \alpha_k + P\{x_k\}$, те је $:\tau'_{-1}\tau'_{-2}\dots\tau'_{-n_{k+1}} = .\tau_{-1}\tau_{-2}\dots\tau_{-n_k} + .\tau''_{-1}\tau''_{-2}\dots\tau''_{-n_k}$. Због $2^{n_i} P\{x_i\} \geq 1, \quad i = \overline{1, \infty}$ лако се види да ће предложене кодне замене за α_k и α_{k+1} бити различити бинарни бројеви што обезбеђује једнозначност декодирања. □

Применити Shannon-ову процедуру кôдирања на следећи скуп порука:

$$[X] = [x_1, x_2, x_3, x_4]$$

$$[P] = [.4, .3, .2, .1].$$

Кренимо редом:

$$\alpha_1 = 0, \quad \alpha_2 = .4, \quad \alpha_3 = .7, \quad \alpha_4 = .9$$

$$.4 \geq 2^{-2}, \quad n_1 = 2$$

$$.3 \geq 2^{-2}, \quad n_1 = 2$$

$$.2 \geq 2^{-3}, \quad n_1 = 3$$

$$.1 \geq 2^{-4}, \quad n_1 = 4$$

$$\begin{array}{ll}
0 = \alpha_1 = 00 & | \quad x_1 \rightarrow 00 \\
.4 = \alpha_2 = 01 & | \quad 1 \quad x_2 \rightarrow 01 \\
.7 = \alpha_3 = 101 & | \quad 1 \quad x_3 \rightarrow 101 \\
.9 = \alpha_4 = 1110 & | \quad x_4 \rightarrow 1110
\end{array}$$

2.6 Gilbert-Moore кôдирање

Кренимо од неједначина:

$$2^{1-n_k} \leq P\{x_k\} < 2^{2-n_k}, \quad k = \overline{1, N}.$$

Да би се уверили да кôд са наведеним дужинама кôдних речи заиста постоји приметимо:

$$\sum_{k=1}^N 2^{1-n_k} \leq 1 < \sum_{k=1}^N 2^{2-n_k}$$

тј.

$$2 \sum_{k=1}^N 2^{-n_k} \leq 1 < 4 \sum_{k=1}^N 2^{-n_k}.$$

дакле, постојање кôда је обезбеђено. Штавише:

$$\begin{array}{l}
1 - n_k \leq \log P\{x_k\} < 2 - n_k \\
1 - \bar{L} \leq -H(X) < 2 - \bar{L} \\
1 + H(X) \leq \bar{L} < 2 + H(X).
\end{array}$$

Алгоритам је следећи:

1. записати поруке у жељеном поретку
2. одабрати $(n_i)_{i=1}^N$ тако да је $2^{1-n_i} \leq P\{x_i\} < 2^{2-n_i}$.
3. израчунати неоппадајући низ:

$$\begin{array}{l}
\alpha_1 = \frac{1}{2}P\{x_1\} \\
\alpha_2 = P\{x_1\} + \frac{1}{2}P\{x_2\} \\
\alpha_3 = P\{x_1\} + P\{x_2\} + \frac{1}{2}P\{x_3\}
\end{array}$$

4. $x_i \rightarrow$ бинарни развој броја α_i на n_i места.

Покажимо да је горњи алгоритам допустив. Једна од следеће две неједнакости мора бити испуњена за било која два симбола ($i < j$):

$$(a) \quad P\{x_i\} \leq P\{x_j\} \quad (b) \quad P\{x_i\} > P\{x_j\}$$

(a) $\Rightarrow n_i \geq n_j$, но, због:

$$\begin{array}{l}
\alpha_j \geq \alpha_i + \frac{1}{2}P\{x_i\} + \frac{1}{2}P\{x_j\} \\
\alpha_j \geq \alpha_i + 2^{-n_i} + 2^{-n_j}
\end{array}$$

добијамо да j -та кôдна реч не може бити једнака са првих n_j места у и-тој кôдној речи. Сличан закључак можемо извести као последицу за (б).

□

Применити Gilbert-Moore-ову процедуру кôдирања на следећи скуп порука:

$$[X] = [\sqcup, A, B, C]$$

$$[P] = [.1859, .0642, .0127, .0218].$$

Решење:

$$[n] = [4, 5, 8, 7] \quad [\alpha] = [.09295, .2180, .25635, .2736]$$

$$\begin{aligned} \sqcup &\rightarrow 0001 \\ A &\rightarrow 00110 \\ B &\rightarrow 01000001 \\ C &\rightarrow 0100011 \end{aligned}$$

Напомена: Слична процедура се може провести и замењујући α_i са $\beta_i = \sum_{j=1}^{i-1} 2^{1-n_j} + 2^{-n_i}$.

2.7 Huffman-ов кôд

Овај кôд се заснива на изградњи дрвета али почевши од лишћа. Када стигнемо до стабла дрвета враћамо се до сваког од листова. Уочимо два најмање вероватна знака и спојимо та два знака у један. Поновимо поступак из претходног корака на новодобијени скуп знакова (са за 1 мањим бројем чланова). Завршавамо када нам остане два члана. Кôдирање сада вршимо тако што тој двојници преосталих дајемо кôдове 0 и 1 (ако је у питању бинарно кôдирање, сасвим је слично за кôдирање азбуком са D слова само што спајамо по D најмање вероватних знакова а не по 2). Сада, у следећем кораку, раздвајамо спојене и сваком од њих додајемо по 0 или 1 на постојећу кôдну реч. Настављамо док нисмо "размотали" све упарене знакове.

Теорема 2.7.1 *Huffman-ов кôд је оптималан.*

Доказ 2.7.1 *Претпоставимо да није оптималан. Нека је C'_1 оптималан кôд за x_1, x_2, x_n . Он ће имати кôдне речи k'_1, k'_2, \dots, k'_n са одговарајућим дужинама d'_1, d'_2, \dots, d'_n .*

Докажимо да за оптималан тренутан кôд важи:

- Знаци са већим вероватноћама имају краће кôдне речи, $p_j > p_k \rightarrow d_j \leq d_k$,
- Два знака са најмањим вероватноћама имају кôдне речи исте дужине, $d_{n-1} = d_n$,

- Међу кЌдним речима дужине k_n мора бити најмање две речи које се слажу на свим местима осим последњег.

Ако је $p - j > p_k$ и $d_j > d_k$ тада је могуће образовати бољи кЌд једноставном заменом речи k_j и k_k .

Нека је $p_{n-1} > p_n$. Тада је $d_{n-1} \leq d_n$. Ако је $p_{n-1} = p_n$ тада, на основу претпоставке о уређењу кодних речи имамо $d_{n-1} \leq d_n$. Напокон, ако је $d_n > d_{n-1}$ можемо одбацити последњи знак n -те кЌдне речи и добити кЌд који је и даље тренутан и бољи од оног од којег смо пошшли.

Ако се никоје две кЌдне речи највеће дужине не поклапају на свим местима осим последњег онда можемо одбацити последњи знак таквих кЌдних речи и добити бољи кЌд.

Дакле, сада имамо да је $d'_{n-1} = d'_n$. Последње две кЌдне речи се поклапају свуда осим на последњем знаку. Без губитка опшности нека су то кодне речи k_{n-1} и k_n . Спојимо x_{n-1} и x_n и образујмо кЌд C'_2 тако што ћемо $x_{n,n-1}$ придружити кЌдну реч k_n са откљоњеним последњим знаком. Проверите да кЌд C'_2 има мању очекивану дужину кодне речи од C''_1 што противуречи претпоставци о оптималности кЌда C''_1 чиме је тврђење доказано. □

2.8 Shanno-Fano-кЌдирање

Једноставно, скуп вероватноћа поделите у два дела (за бинарно кЌдирање, на D делова ако је азбука обима D) што приближније једнаких збирова. Наставите тако са дељењем подскупова све док не постану двочлани. Сваки од чланова последњих добијених двочланих скупова кЌдирати са 0 и 1, и наставити са "пењањем" назад, додајући сваки пут по један (0 или 1) знак код "распакивања" сваког од подскупова.

Ако вам се ово учинило познатим, то је управо правило које смо предложили на почетку када смо уводили појам неодређености. Тиме смо затворили тај круг.

У овом поглављу смо показали како припремити податке за пренос преко неког од комуникационих канала везе. Наиме, отклонили смо сувишност својствену сваком од извора података. На основу те сувишности често смо способни да и на основу делимичног садржаја неке поруке реконструишемо целу поруку. Пример је отклањање самогласника из речи. Управо то је идеја следећег поглавља где ћемо се бавити како да сажете податке поуздано пренесемо преко задатог канала везе.

Останите уз нас, не мењајте канал!

Поглавље 3

Канали везе

3.1 Дискретан канал везе без памћења

Канал везе ћемо задати као математички модел за, поготову данас, веома чест, скоро неизбежан, задатак преноса података преко неког од медија, тј. начина остваривања везе. Била то радио веза, жицна веза у ма ком облику или нешто треће све ћемо то покушати да опишемо једним математичким моделом.

Као прво, задајмо улазну и излазну азбуку тј. скупове знакова које можемо очекивати на улазу тј. излазу из канала везе. Означимо их са \mathcal{U} тј. \mathcal{V} . За потребе овог поглавља (отуд појам "дискретан" канал везе) ограничићемо се на коначне улазну и излазну азбуку. Случајно променљиве које очекујемо на улазу тј. излазу из канала везе означимо са U тј. V . Одмах да приметимо да, супротно свим очекивањима, нисмо означили улаз у канал са X јер ћемо ту ознаку користити за случајно променљиву која представља извор података који желимо да пренесемо каналом везе. Између X и U наћи ће се кодирање без шума задужено да отклони сувишност, као што смо то обрадили у претходном поглављу.

Основно ограничење на модел канала везе јесте постојање шума у преносу података током остваривања везе. Шум ћемо, као што је уобичајено у применама теорије вероватноће, моделирати случајним процесом. Но, за разлику од уобичајених метода, шум нећемо издвајати и описивати његову расподелу вероватноће или нешто слично, већ ћемо, једноставно, посматрати збирку условних вероватносних мера $\{p(v_j|u_i); i \in \mathcal{U}, j \in \mathcal{V}\}$. Дакле, за свако улазно слово из улазне азбуке имамо условну расподелу слова из излазне азбуке. Случајно променљива Y је, тиме, сасвим добро задата.

На тај начин, чим задамо расподелу вероватноће случајно променљиве U , имамо задате све следеће величине: $H(U)$, $H(V)$, $H(V|U)$, $H(U|V)$ и $I(X; Y)$.

Запазимо да је канал везе сасвим одређен само збирком условних вероватноћа, док, за познавање расподеле вероватноће на његовом излазу неопходно је да знамо и расподелу вероватноће на улазу.

Једна од основних нумеричких особина канала везе јесте и, у неком

смислу, мера његове пропусне моћи тј. **капацитет** канала везе који задајемо као:

$$\mathfrak{C} = \sup_{\mathbb{P}_U} I(U; V)$$

где је \mathbb{P}_U расподела вероватноће случајно променљиве U .

Један од осовних задатака теорије информације јесте доказивање онога што је Shannon назвао теоремом кодирања за канал везе. Дакле, ваља наћи услове које мора задовољити канал везе да би могли тврдити да постоји начин да или скоро сасвим поуздано (са вероватноћом 1) или са неком великом вероватноћом пренесемо податке тим каналом везе. Већ у доказивању Фапо-ове неједнакости имали смо појам грешке јер је та неједнакост и доказивана управо ради употребе у доказу теореме кодирања за канал везе. Још једна неједнакост коју смо већ доказали а прављена је по мери канала везе као модела јесте Data Processing Inequality.

Остаје да објаснимо и откуд "без памћења" у наслову овог одељка. Дакле, као што можете закључити из горњег увода, вероватноћа сваког излазног слова из канала везе зависи искључиво од слова које је управо примљено на улазу. Отуд недостатак памћења. Наравно, постоје знатно општији модели канала везе.

Пре него што кренемо даље, уочимо да се код дискретног канала без памћења, када говоримо о збирци условних расподела, у ствари ради о једноставној матрици условних вероватноћа где су врсте обележене улазном азбуком канала везе, а колоне излазном азбуком канала везе.

3.2 Врсте дискретних канала везе без памћења

Отпратићемо излагање Ash-а у овом разврставању и развијању особина информације као функције матрице прелазних вероватноћа и улазно-излазних расподела вероватноће.

Прво разврставање дискретних канала везе без памћења извршићемо по $I(U; V)$.

- За канал везе кажемо да је без губитака ако је $H(U|X) = 0$. Тада је улаз у канал везе сасвим одређен његовим излазом. Дакле, скуп вредности излаза канала везе можемо поделити у дисјунктне скупове

$$E_1, E_2, \dots, E_N \subset \mathcal{V} \text{ такве да је } \mathbb{P}\{V \in E_i | V = v_i\} = 1, \quad i = \overline{1, N}.$$

- За канал везе кажемо да је неслучајан ако

$$\forall i, j \quad p(v_j | u_i) = 1 \text{ или } 0$$

тј. ако је (еквивалентно) $H(V|U) = 0$.

- За канал везе кажемо да је без шума ако је и без губитака и неслучајан.

- За канал везе кажемо да је бескорисан ако је $I(U|V) = 0$ тј. ако је (еквивалентно) $H(U|V) = H(U)$, тј. ако су (еквивалентно) U и V независне случајно променљиве, за све улазне расподеле вероватноће.
- За канал везе кажемо да је симетричан ако су скупови вероватноћа сви врста међусобно једнаки и скупови вероватноћа свих колона такође међусобно једнаки.

Код симетричног дискретног канала везе без памћења лако доказујемо да је $H(V|U)$ независно од улазне расподеле вероватноће и да зависи искључиво од матрице условних вероватноћа.

$$\begin{aligned} H(V|U = u_i) &= - \sum_{j=1}^{\#\mathcal{V}} p'_j \log p'_j, \quad i = \overline{1, \#\mathcal{U}} \\ H(V|U) &= - \sum_{i=1}^{\#\mathcal{U}} p(u_i) H(V|U = u_i) \\ &= - \sum_{j=1}^{\#\mathcal{V}} p'_j \log p'_j, \end{aligned}$$

за сваку улазну расподелу вероватноћа $p(u)$. Један од важних примера симетричног канала везе јесте бинаран симетричан канал везе без памћења. Дакле, случај када је $\#\mathcal{U} = \#\mathcal{V} = 2$. Из симетричности лако закључујемо да је, скраћено записано, $p(0,0) = p(1,1) = 1 - \epsilon$ и $p(0,1) = p(1,0) = \epsilon$.

Вратимо се на општи симетричан канал везе без памћења. користећи се изразом који смо извели горе и стављајући да је

$$p(u_i) = \frac{1}{\#\mathcal{U}}, \quad i = \overline{1, \#\mathcal{U}}.$$

Рачунајући добијамо

$$p(v_j) = \sum_{i=1}^{\#\mathcal{U}} p(u_i, v_j) = \sum_{i=1}^{\#\mathcal{U}} p(u_i) p(v_j|u_i) = \frac{1}{\#\mathcal{U}} p(v_j|u_i).$$

Последњи збир јесте збир j -те колоне у матрици тог канала везе. Та матрица је симетрична те имамо да је

$$\sum_{i=1}^{\#\mathcal{U}} p(v_j|u_i) = \sum_{l=1}^{\#\mathcal{U}} s'_l, \quad \text{независно по } j.$$

Дакле, све вероватноће у расподели на излазу из канала везе су једнаке те је, када уврстимо $H(Y) = \log \#\mathcal{V}$ у израз за информацију, користећи већ израчунату вредност $H(V|U)$,

$$\mathfrak{C} = \log \#\mathcal{V} - H(p'_1, \dots, p'_{\#\mathcal{V}}).$$

3.3 Метода Muroga

Израчунаћемо капацитет канала задатог матрицом прелазних вероватноћа. Ограничење на ту матрицу јесте да она мора бити квадратна тј. да је $\#\mathcal{U} = \#\mathcal{V}$.

За дату матрица прелазних вероватноћа $(p_{ij})_{m \times m}$ имамо да је:

$$H(V | U) = - \sum_{i=1}^m p(u_i) \sum_{j=1}^m p(v_j | u_i) \log(p(v_j | u_i)).$$

Напишимо следећи систем:

$$\begin{aligned} p_{11}A_1 + \dots + p_{1m}A_m &= \sum_{j=1}^m p(v_j | u_1) \log(p(v_j | u_1)) = \sum_{j=1}^m p_{1j} \log(p_{1j}) \\ &\dots \\ p_{m1}A_1 + \dots + p_{mm}A_m &= \sum_{j=1}^m p(v_j | u_m) \log(p(v_j | u_m)) = \sum_{j=1}^m p_{mj} \log(p_{mj}). \end{aligned}$$

Одатле можемо записати нашу условну неодређеност као:

$$\begin{aligned} H(V | U) &= -p_1(p_{11}A_1 + \dots + p_{1m}A_m) - p_2(p_{21}A_1 + \dots \\ &\quad + p_{2m}A_m) - \dots - p_m(p_{m1}A_1 + \dots + p_{mm}A_m). \end{aligned}$$

Означимо:

$$p_1 = p(u_1) \dots p_i = p(u_i).$$

Приметимо да је:

$$p(v_1) = p_1 p_{11} + \dots + p_m p_{m1} =: q_1$$

...

$$p(v_m) = p_1 p_{1m} + \dots + p_m p_{mm} =: q_m.$$

Одатле имамо да је:

$$H(V | U) = -q_1 A_1 - \dots - q_m A_m.$$

Закључујемо даље:

$$I(U; V) = H(V) - H(V | U) = - \sum_{i=1}^m q_i \log(q_i) + \sum_{i=1}^m q_i A_i.$$

Даље се све одвија Лагранге-овом методом множилаца за налажење екстремума функције више променљивих.

Дефинишимо функцију:

$$U(q_1, \dots, q_m) = - \sum_{i=1}^m q_i \log(q_i) + \sum_{i=1}^m q_i A_i + \lambda \sum_{i=1}^m q_i,$$

налазећи изводе добијамо:

$$\frac{\partial U}{\partial q_i} = -(1 + \log(q_i)) + A_i + \lambda = 0$$

тј.

$$-1 + A_1 - \log(q_1) + \lambda = 0$$

...

$$-1 + A_m - \log(q_m) + \lambda = 0,$$

одакле, због истовремености, добијамо:

$$i \neq j \Rightarrow U_i - \log(q_i) = U_j - \log(q_j),$$

помножимо сваку од веза са q_i и применимо оператор $\sum_{i=1}^m$:

$$-\sum_{i=1}^m q_i + \sum_{i=1}^m q_i (A_i - \log(q_i)) + \lambda \sum_{i=1}^m q_i = 0$$

тј.

$$-1 + A_i - \log(q_i) + \lambda = 0$$

или

$$A_i - \log(q_i) = 1 - \lambda.$$

Дакле,

$$I(U; V) = \sum_{i=1}^m q_i (A_i - \log(q_i)) = \sum_{i=1}^m q_i (1 - \lambda) = 1 - \lambda = C.$$

Даље, непосредно имамо:

$$C = 1 - \lambda = A_i - \log(q_i), \quad q_i = 2^{A_i - C}$$

што, због чињенице да су ч-ови вероватноће и у збиру дају један закључујемо:

$$C = \log(2^{A_1} + \dots + 2^{A_m}).$$

3.4 Метода Meister, B. & Oetti, W. 1957

Ево услова који су неопходни да би ова метода довела до капацитета канала везе са задатом матрицом прелазних вероватноћа:

- матрица не мора бити квадратна и не мора бити несингуларна,
- води се рачуна и о цени преноса сваког знака,
- метода се заснива на конвексности фје неодређености и проблем решавамо методама квази-конвексног програмирања.

ПРОБЛЕМ:

Дат нам је дискретан канал везе без меморије својом матрицом прелазних вероватноћа:

$$P = (p_{ij}), i = \overrightarrow{1, n}, j = \overrightarrow{1, m}$$

$$p_{ij} \geq 0, i = \overline{1, n}, j = \overline{1, m}, \sum_{i=1}^n p_{ij} = 1, j = \overline{1, m}$$

Претпоставимо, даље, да нема нултих врста. Дакле, за сваку расподелу на улазу:

$$\vec{p} = (p_1, \dots, p_n); p_i \geq 0, i = \overline{1, n}, \sum_{i=1}^n p_i = 1.$$

и имамо задати "ценовник" улазних знакова:

$$\vec{c} = (c_1, \dots, c_n); c_i \geq 0, i = \overline{1, n}.$$

Брзина преноса кроз дати канал везе се дефинише као:

$$R(p) := \sum_{i=1}^n \sum_{j=1}^m p_j p_{ij} \ln \left(\frac{p_{ij}}{\sum_{k=1}^m p_k p_{ik}} \right) \frac{1}{\sum_{j=1}^m c_j p_j}.$$

Дакле:

$$C = \max_{\vec{p} \in \mathcal{P}} R(p), \quad \mathcal{P} : \text{скуп допустивих расподела } \vec{p}.$$

За наше потребе је довољно узети:

$$\mathcal{P} = \{ \vec{p} \mid p_i \geq 0, i = \overline{1, n}; \sum_{i=1}^n p_i = 1 \}.$$

Уведимо следеће ознаке:

$$\alpha_j := \sum_{i=1}^n p_{ij} \ln(p_{ij})$$

$$q_i := \sum_{j=1}^m p_{ij} p_j,$$

(излазна расподела вероватноћа)

$$(\ln(\vec{q}))_i = \ln(q_i), i = \overline{1, m}$$

$$\vec{s} = (p_1, \dots, p_n, q_1, \dots, q_m)$$

$$S = \{ \vec{s} \mid p_j \geq 0, \sum_{j=1}^n p_j = 1, q_i = \sum_{j=1}^m p_{ij} p_j \}.$$

$$R(s) = \frac{\langle \vec{\alpha}, \vec{p} \rangle - \langle \vec{q}, \ln(\vec{q}) \rangle}{\langle \vec{c}, \vec{p} \rangle} =: \frac{f(\vec{s})}{g(\vec{s})}.$$

Проблем, сада, постаје:

$$C = \max_{\vec{s} \in S} R(\vec{s})$$

Морамо проверити:

- S је конвексан полиедар у E^{n+m}
- f и g су позитивне над S
- R је непрекидна у S и непрекидно диференцијабилна у непразном скупу:

$$S^0 := \{\vec{s} \mid s \in S, q_i > 0, i = \overline{1, m}\}.$$

Дакле, R достиже свој максимум за барем једно $\vec{s} \in S$.

Пређимо на методу:

корак 1 : почињемо са $\vec{s} \in S^0$ произвољним (обично се креће са равномерном расподелом)

корак 2 : за дато $\vec{s}^k \in S^0$ налазимо $\vec{s}^k \in S$ тако да је:

$$r_{S^k} = \max_{\vec{s} \in S} r_{S^k}(\vec{s})$$

$$r_{S^1}(\vec{s}) = \frac{f(\vec{s}^1) + \langle \partial f(\vec{s}^1), (\vec{s} - \vec{s}^1) \rangle}{g(\vec{s}^1) + \langle \partial g(\vec{s}^1), (\vec{s} - \vec{s}^1) \rangle}$$

корак 3 : на одсечку

$$[\vec{s}^k, \vec{s}^k]$$

одређујемо тачку \vec{s}^{k+1} , такву да је:

$$R(\vec{s}^{k+1}) = \max_{\vec{s} \in [\vec{s}^k, \vec{s}^k]} R(s)$$

корак 4 : вратити се на корак 2 са вредношћу \vec{s}^{k+1} .

Веома успешну методу за одређивање капацитета канала можете наћи у раду Suguru Arimoto, "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels", IEEE Transactions on Information Theory, vol. IT-18, No. 1, January 1972.

Поглавље 4

Пренос података каналом везе

4.1 О кôду и грешци

Претпоставићемо да смо податке припремили за пренос каналом везе, тј. да смо отклонили сваку могућу сувишност као што смо то објаснили раније (теорема кодирања за канале везе без шума).

Поруку коју шаљемо бирамо из уређеног скупа $W \in \{1, 2, \dots, M\}$ добијајући $u^m(W)$. Дакле, не шаљемо слово по слово кроз канал везе већ, шодно процени, слова повезујемо у m -торке (ниске).

Након проласка ниске кроз канал везе, на излазу, добијамо случајну ниску v^m сагласно условној расподели вероватноћа $\mathbf{p}(v^m|u^m)$ где је $\mathbf{p}(v_k|u^k, v^{k-1}) = \mathbf{p}(v_k|u_k)$, $k = 1, 2, \dots, m$ што у случају да је канал без повратне спреге, тј. ако улазни знакови не зависе од прошлих излазних знакова, тј. $\mathbf{p}(u_k|u^{k-1}, v^{k-1})$, даје $\mathbf{p}(v^m|u^m) = \prod_{i=1}^m \mathbf{p}(v_i|u_i)$.

Сагласно одговарајућем правилу препознавања погађамо обележје за W . Ако се обележје не поклапа учинили смо грешку препознавања, у супротном смо правилно препознали.

(M, m) -кôд за канал $(\mathcal{U}, \mathbf{p}(v|u), \mathcal{V})$ састоји се од:

- Уређеног скупа $\{1, 2, \dots, M\}$.
- Пресликавања (кôдирања) $u^m : \{1, 2, \dots, M\} \rightarrow \mathcal{U}^m$ које даје кôдне речи $u^m(1), u^m(2), \dots, u^m(M)$. Скуп кôдних речи називамо кôдном књигом.
- Пресликавања препознавања (декôдирања) $g : \mathcal{V}^m \rightarrow \{1, 2, \dots, M\}$, које је неслучајно правило које свакој могућој излазној n -торци придружује оцену.

Нека је:

$$\lambda_i = \mathbb{P}\{g(v^m) \neq i | u^m = u^m(i)\} = \sum_{v^m} \mathbf{p}(v^m|u^m(i)) \mathbf{1}_{g(v^m) \neq i}$$

условна вероватноћа грешке ако је послата реч са обележјем i , где је $\mathbf{1}_A$ индикатор догађаја A .

Највећа вероватноћа грешке за (M, m) -кôд је

$$\lambda^{(m)} = \max\{\lambda_i | i = 1, 2, \dots, M\}.$$

Средња вероватноћа грешке за (M, m) -кôд је

$$\mathbf{P}_e^{(m)} = \frac{1}{M} \sum_{i=1}^M \lambda_i.$$

Брзина (M, m) -кôда је $\mathbf{R} = \frac{\log M}{m}$ бита по сваком појединачном преносу.

Брзина преноса \mathbf{R} је достижна ако постоји низ $(\lceil 2^{m\mathbf{R}} \rceil, m)$ кôдова таквих да највећа вероватноћа грешке $\lambda^{(m)} \rightarrow 0$, $m \rightarrow \infty$.

Капацитет канала везе је горња међа (supremum) свих достижних брзина преноса.

Наслућујемо да ће, за довољно дуге кôдне речи, при брзинама непосредно испод капацитета бити могуће остварити произвољно малу вероватноћу грешке.

Свакако, постоје разни приступи правилима препознавања. Међутим, скоро сва користе особину коју називамо АЕР (Asymptotic Equipartition Property). Не, нећу покушавати да то преведем.

Ако посматрамо на ту особину из теорије вероватноћа то је веома једноставна последица закона великих бројева. Закон великих бројева каже да ако имамо низ независних једнако расподељених случајно променљивих X_i , $i = 1, 2, \dots$ тада $\frac{1}{n} \sum_{i=1}^n X_i$ је за довољно велико n блиско $\mathbb{E}X_1$.

Теорема 4.1.1 (АЕР) *Нека су X_i , $i = 1, 2, \dots$ независне и једнако расподељене (сагласно $\mathbf{p}(x)$). Тада $-\frac{1}{n} \log \mathbf{p}(X_1, X_2, \dots, X_n) \rightarrow H(X_1)$ у вероватноћи.*

Доказ 4.1.1 *Ако применимо исто пресликавање на независне једнако расподељене случајно променљиве добијамо независне једнако расподељене случајно променљиве. У нашем случају нека је то пресликавање $-\log \mathbf{p}(X_i)$. Тада, сходно закону великих бројева, имамо:*

$$\begin{aligned} -\frac{1}{n} \log \mathbf{p}(X_1, X_2, \dots, X_n) &= -\frac{1}{n} \sum_{i=1}^n \log \mathbf{p}(X_i) \\ &\rightarrow -\mathbb{E} \log \mathbf{p}(X_1) \text{ у вероватноћи} \\ &= H(X_1) \end{aligned}$$

□

Ова једноставна последица је веома значајна јер нам омогућује да скуп n -торки поделимо на два подскупа. Један, скуп типичних, приближно једнаких вероватноћа које у збиру износе блиско јединици и, други, скуп нетипичних који у збиру имају вероватноћу блиску нули.

За ниску (x_1, x_2, \dots, x_n) ћемо рећи да је типична (припада скупу \mathcal{T}_ϵ^n) ако је

$$2^{-n(H(X_1)+\epsilon)} \leq \mathbf{p}(x_1, x_2, \dots, x_n) \leq 2^{-n(H(X_1)-\epsilon)}.$$

Теорема 4.1.2 *Скуп \mathcal{T}_ϵ^n има следеће особине:*

1. Ако је $(x_1, x_2, \dots, x_n) \in \mathcal{T}_\epsilon^n$ тада је

$$H(X_1) - \epsilon \leq -\frac{1}{n} \log \mathbf{p}(x_1, x_2, \dots, x_n) \leq H(X_1) + \epsilon.$$

2. $\mathbb{P}\{\mathcal{T}_\epsilon^n\} > 1 - \epsilon$ за n довољно велико.

3. $\#\mathcal{T}_\epsilon^n \leq 2^{n(H(X_1)+\epsilon)}$.

4. $\#\mathcal{T}_\epsilon^n \geq 2^{n(H(X_1)-\epsilon)}$ за n довољно велико.

Доказ 4.1.2 за 1. Следи непосредно на основу дефиниције скупа \mathcal{T}_ϵ^n .

за 2. Следи непосредно на основу претходно доказане теореме. Наиме, вероватноћа догађаја $(X_1, X_2, \dots, X_n) \in \mathcal{T}_\epsilon^n$ тежи 1 за $n \rightarrow \infty$. Дакле, за свако $\delta > 0$ постоји n_0 такво да је за све $n \geq n_0$ испуњено

$$\mathbb{P}\left\{\left| -\frac{1}{n} \log \mathbf{p}(X_1, X_2, \dots, X_n) - H(X_1) \right| < \epsilon\right\} > 1 - \delta.$$

Стављајући да је $\delta = \epsilon$ добијамо 2.

за 3.

$$\begin{aligned} 1 &= \sum_{\mathbf{x} \in \mathcal{X}^n} \mathbf{p}(\mathbf{x}) \\ &\geq \sum_{\mathbf{x} \in \mathcal{T}_\epsilon^n} \mathbf{p}(\mathbf{x}) \\ &\geq \sum_{\mathbf{x} \in \mathcal{T}_\epsilon^n} 2^{-n(H(X_1)+\epsilon)} \\ &= 2^{-n(H(X_1)+\epsilon)} \#\mathcal{T}_\epsilon^n, \text{ тј.} \\ \#\mathcal{T}_\epsilon^n &\leq 2^{n(H(X_1)+\epsilon)}. \end{aligned}$$

за 4. За довољно велико n имамо да је $\mathbb{P}\{\mathcal{T}_\epsilon^n\} > 1 - \epsilon$ те је

$$\begin{aligned} 1 - \epsilon &< \mathbb{P}\{\mathcal{T}_\epsilon^n\} \\ &\leq \sum_{\mathbf{x} \in \mathcal{T}_\epsilon^n} 2^{-n(H(X_1)-\epsilon)} \\ &= 2^{-n(H(X_1)-\epsilon)} \#\mathcal{T}_\epsilon^n, \text{ тј.} \\ \#\mathcal{T}_\epsilon^n &\geq (1 - \epsilon) 2^{n(H(X_1)-\epsilon)}. \end{aligned}$$

□

Пре него са упутимо ка главној теореме тероје информације обратимо пажњу на тврђење које нам у напредним (најопштијим) случајевима доказа горепоменуте теореме значајно помаже.

Укупна вероватноћа грешке у преносу дата је као $\sum_i \mathbf{P}(i)\lambda_i$ где је \mathbf{P} расподела вероватноће на улазу у канал везе.

Теорема 4.1.3 *Нека нам је дат канал везе. Нека је $\lambda^{(m)}(\mathbf{p}, \mathbf{P})$ вероватноћа грешке ако је расподела шума у каналу задата са \mathbf{p} а расподела на улазу дата са \mathbf{P} . Нека је, даље, \mathbf{P}^0 равномерна расподела вероватноће на улазу у дати канал везе. Тада је*

$$\lambda^{(m)}(\mathbf{p}, \mathbf{P}) \leq \lambda^{(m)}(\mathbf{p}, \mathbf{P}^0).$$

Доказ 4.1.3 *Задајмо пресликавања кôдирања и препознавања f и g . Нека је $\lambda(u) = \sum_{y:g(y) \neq (u)} \mathbf{p}(y|f(u))$. Дакле, то је вероватноћа грешке ако је послато u . Укупна вероватноћа грешке је, тада, $\sum_{u \in \mathcal{U}} \mathbf{P}(u)\lambda(u)$. Ако променимо редослед кôдних речи (пермутацијом π) добијемо нову укупну вероватноћу грешке $\lambda(\pi) = \sum_{u \in \mathcal{U}} \mathbf{P}(u)\lambda(\pi(u))$. У случају када је $\mathbf{P} = \mathbf{P}^0$ укупна вероватноћа грешке не зависи од π и једнака је $\frac{1}{\#\mathcal{U}} \sum_{u \in \mathcal{U}} \lambda(u) = \bar{\lambda}$.*

За сваку расподелу вероватноћа \mathbf{P} постоји π таква да је $\lambda(\pi) \leq \bar{\lambda}$. Једноставно, узмимо случајну пермутацију Π , где су све пермутације скупа \mathcal{U} једнако вероватне. Тада је

$$\min_{\pi} \lambda(\pi) \leq \mathbb{E}\lambda(\pi) = \mathbb{E} \sum_{u \in \mathcal{U}} \mathbf{P}(u)\lambda(\pi u)$$

што је једнако

$$\sum_{u \in \mathcal{U}} \mathbf{P}(u)\mathbb{E}\lambda(\pi u) = \sum_{u \in \mathcal{U}} \mathbf{P}(u) \frac{1}{\#\mathcal{U}} \sum_{\tilde{u} \in \mathcal{U}} \lambda(\tilde{u}) = \bar{\lambda}.$$

Дакле, за свако f и g можемо наћи нова правила кôдирања и препознавања са укупном вероватноћом грешке не већом од $\lambda(\mathbf{P}^0, f, g)$. Минимизујући по f и g добијемо тврђење. \square

4.2 Правила препознавања

У општем случају правила препознавања се деле на основу тога да ли онај који врши препознавање познаје расподелу вероватноће на улазу у канал везе или не. У првом случају кажемо да имамо идеалног посматрача док у другом случају имамо препознавање по правилу највеће веродостојности.

У правилу **идеалног посматрача** препознајемо послату реч на основу примљене тражећи највећу накнадну (а posterior-ну) вероватноћу.

$$\mathbb{P}\{\text{пслато је } u \mid \text{примљено је } v\} = \frac{\mathbf{P}(u)\mathbf{p}(v|u)}{p(v)},$$

где је

$$p(v) = \sum_{\mathbf{u}} \mathbf{P}(\mathbf{u})\mathbf{P}(v|\mathbf{u}).$$

У правилу **највеће веродостојности** препознајемо послату реч на основу примљене тражећи највећу претходну (а ригор-ну) вероватноћу $\mathbf{p}(v|u)$.

Теорема 4.2.1 (i) *За свако правило препознавања, идеални посматрач даје најмању укупну вероватноћу грешке међу свим препознавачима.*

(ii) *Ако је улазна порука равномерно расподељена над скупом \mathcal{U} и ако је правило кодирања такво да су све могуће кодне речи за сваку поруку једнако вероватне из скупа свих кодних речи тада се правило идеалног посматрача и правило највеће веродостојности поклапају.*

Доказ 4.2.1 *Пошто је именилац задат примљеном речју v остаје нам да бројилац учинимо што је могуће већим. Претпоставимо да користимо функције f и g за кодирање тј. препознавање. Тада је укупна вероватноћа грешке*

$$\begin{aligned} & \sum_U \mathbf{P}(W = i) \mathbf{p}(g(v) \neq u | \text{ послато је } u(i)) \\ &= \sum_u \sum_{y: g(y) \neq u} \mathbf{p}(y|u) \\ &= \sum_y \sum_{x: x \neq g(y)} p(x) \mathbf{p}(y|x) \\ &= \sum_y \sum_{\text{чч сви } x} p(x) \mathbf{p}(y|x) - \sum_y p(g(y)) \mathbf{p}(y|g(y)) \\ &= 1 - \sum_y p(g(y)) \mathbf{p}(y|g(y)). \end{aligned}$$

Управо правило идеалног посматрача тражи да сваки члан последњег збира буде највећи могући. Дакле, цео збир је, тада, највећи а укупна вероватноћа грешке најмања.

Други део теореме се доказује једноставном рачуницом. □

Теорема 4.2.2 *Ако су изворне поруке једнако расподељене над скупом \mathcal{U} и ако користимо правило највеће веродостојности и кодирамо правилном f , тада укупна вероватноћа грешке задовољава*

$$\lambda(f) \leq \frac{1}{\#\mathcal{U}} \sum_{u \in \mathcal{U}} \sum_{u' \in \mathcal{U}: u' \neq u} \mathbb{P}\{\mathbf{p}(Y|f(u')) \geq \mathbf{p}(Y|f(u)) | U = u\}.$$

Доказ 4.2.2 *Могућности за грешку су:*

- $\mathbf{p}(Y|f(u')) \geq \mathbf{p}(Y|f(u))$ за неко $u' \neq u$,

- $\mathbf{p}(Y|f(u')) = \mathbf{p}(Y|f(u))$ за неко $u' \neq u$ што укључује и случај када је $f(u) = f(u')$,

тј. да нема грешке када је $\mathbf{p}(Y|f(u')) < \mathbf{p}(Y|f(u))$, за свако $u' \neq u$.
Тада је вероватноћа грешке

$$\begin{aligned} & \mathbb{P}(\text{учињена је грешка} | U = u) \\ & \leq \mathbb{P}\{\mathbf{p}(Y|f(u')) \geq \mathbf{p}(Y|f(u)), \text{ за неко } u' \neq u | U = u\} \\ & \leq \sum_{u' \in \mathcal{U}: u' \neq u} \mathbb{P}\{\mathbf{p}(Y|f(u')) \geq \mathbf{p}(Y|f(u)) | U = u\}. \end{aligned}$$

Множећи са $\frac{1}{\#\mathcal{U}}$ и сабирајући по u завршавамо доказ. □

Тврђење се лако преправља у случају да није у питању равномерна расподела већ нека друга, задата (рецимо q), заменом множиоца $\frac{1}{\#\mathcal{U}}$ са $q(u)$.

Поред неслучајних правила кодирања и препознавања користимо и случајна правила кодирања где се кодна реч бира случајно из кодне књиге. Предности случајног кодирања су то што из постојања ”доброг” случајног кодера следи и постојање ”доброг” неслучајног и што се рачунање код случајних кодова уместо дискретне оптимизације своди на оптимизацију над вероватносним расподелама што је удобније. Мана је та што, иако са занемарљивом вероватноћом може доћи до случаја да добијени код није један-у-један, тј да се кодне речи за две различите поруке поклопе.