Jasna Prezelj

Verjetnost in statistika

za kemike, 2. stopnja

Ljubljana 2012

Kazalo

1. Uvodni primeri

1. V 17. stoletju je bila med italijanskimi kockarji popularna stava na skupno število pik na treh kockah. Možni je bilo staviti na dogodek, da bo skupno število pik enako določeni številki. Če je bil seštevek pik po metu enak tistemu, na katerega je igralec stavil, je le-ta stavo dobil, sicer pa izgubil. Najpogostejši stavi sta bili na vsoto 9 ali 10. Po razmišljanju takratnih kockarjev sta bili obe stavi enako verjetni, saj je kombinacij števil, ko dajo vsoto 9 ali 10 v obeh primerih enako.

vsota 9	vsota 10
1 2 6	1 3 6
1 3 5	1 4 5
1 4 4	2 2 6
2 2 5	2 3 5
2 3 4	2 4 4
3 3 3	3 3 4

Kockarska praksa pa je pokazala, da se število 10 pojavi večkrat kot 9. Da bi problem rešili, so se obrnili na slavnega sodobnika Galilea (1564 - 1642). Ta je problem rešil tako, da je izpisal vse možne izide metov treh kock in preštel število izidov za posamezno vsoto. Upošteval je, da sta pri metu treh kock npr. izida 1 2 6 in 2 1 6 različna. Tako kombinacijo števil lahko dobimo v šestih primerih:

$$(1,2,6),(1,6,2),(2,1,6),(2,6,1),(6,1,2),(6,2,1).$$

Zapišimo vse možne vrstne rede števil:

vsota 9	št. vrst. red.	vsota 10	št. vrst. red.
1 2 6	6	1 3 6	6
1 3 5	6	1 4 5	6
1 4 4	3	2 2 6	3
2 2 5	3	2 3 5	6
2 3 4	6	2 4 4	3
3 3 3	1	3 3 4	3

Za vsoto 9 je ugodnih izidov 25, za vsoto 10 pa 27. Vseh možnih izidov je $6 \cdot 6 \cdot 6 = 216$, torej je verjetnost, da pade 9 enaka 25/216, verjetnost, da pade 10 pa 27/216.

Za izračun ugodnih izidov pri kockanju je boli potrebno poznati število vseh možnih razporeditev danih treh števil in število vseh možnih izidov.

Trditev 1.2. *Dana so števila od* 1 *do* n. *Posamezno razvrstitev teh števil imenujemo* **permutacija** n *števil. Število vseh (različnih) permutacij* n *števil je* $n!$.

Dokaz. Predstavljajmo si, da imamo oštevilčene kroglice, ki jih razvrščamo v n predalčkov. Na slepo izberemo prvo kroglico in jo damo v prvi predal. Možnih izborov je n. Drugo kroglico bomo izbrali izmed preostalih $n-1$ kroglic, zato imamo za drugo mesto $n-1$ možnosti. Za tretje mesto imamo na voljo še $n-2$ kroglic, torej imamo $n-2$ možnosti itd. Skupno število vseh možnih permutacij je $n(n-1)(n-2)\cdot\ldots\cdot 3\cdot 2\cdot 1 = n!$. ♢

Na koliko načinov lahko števila od 1 do n razporedimo na r mest ($r \leq n$)? Razvrstitev n različnih števil na r mest imenujemo **variacija reda** r **med** n **elementi brez ponavljanja**. Število vseh možnih variacij reda r med n elementi označimo s V_n^r. Za izračun V_n^r razmišljamo podobno kor zgoraj. Za prvo mesto imamo n možnosti, za drugo $n-1$ itd. Za mesto $r-1$ imamo $n-(r-2)$ možnosti in za r-to mesto $n-r+1$, skupaj $n(n-1)\cdot\ldots\cdot(n-r+2)(n-r+1)$,

$$V_n^r = n(n-1)\cdot\ldots\cdot(n-r+2)(n-r+1) = \frac{n!}{(n-r)!}.$$

Kaj pa, če nas vrstni red elementov ne zanima, ampak hočemo vedeti le, koliko je različnih podmnožic z r elementi v množici z n elementi? Podmnožico množice z n elementi, ki ima r elementov ($r \leq n$) imenujemo **kombinacija reda** r **med** n **elementi**. Število vseh takih kombinacij iznačimo z C_n^r. Za izračun si pomagamo z variacijami V_n^r. Ker smo pri variacijah upoštevali vrstni red, bomo iz vsake kombinacije r števil dobili $r!$ variacij, zato je

$$C_n^r = \frac{V_n^r}{r!} = \frac{n!}{(n-r)!r!} = \left(\begin{array}{c} n \\ r \end{array}\right).$$

Variacija reda r **med** n **elementi s ponavljanjem** razporeditev n števil v r predalčkov, če dopuščamo, da se števila ponavljajo. To pomeni, da izmed n kroglic v škatli, oštevilčenih od 1 do n izberemo prvo, število zapišemo v

prvi predal, kroglico vrnemo v škatlo in ponovimo izbiranje. V tem primeru imamo za vsakega od predalov n možnosti, predalov je r, kar da skupaj n^r množnosti. Število variacij reda r med n elementi s ponavljanjem je enako $V_n^{r\,(p)} = n^r$.

Na podoben način so definirane kombinacije brez ponavljanja. Imamo posodo z n kroglicami, ki so oštevilčene od 1 do n. Radi bi sestavili množico z r elementi tako: izberemo prvo kroglico, njeno številko napišemo na listek in ga damo v škatlo, kroglico pa vrnemo v posodo. Postopek r-krat ponovimo. Dobljena množica števil je **kombinacija reda r med n elementi s ponavljanjem**. Število vseh takih kombinacij označimo s $C_n^{r\,(p)}$, izračunamo pa ga tako. Uredimo števila v škatli po velikosti od najmanjšega do največjega in sestavimo novo zaporedje števil po naslednjem postopku: prvo število pustimo, drugemu prištejemo 1, tretjemu 2, in tako dalje. Zadnjemu številu smo prišteli $r-1$. Dobili smo urejeno množico r števil, ki pa so po velikosti lahko med 1 in $n+r-1$. Izberimo si zdaj poljubno podmnožico z r elementi izmed množice števil z $n+r-1$ elementi in jo uredimo po velikosti. Če prvo število pustimo, drugemu odštejemo 1, tretjemu 2 itd, bomo dobili kombinacijo s ponavljanjem reda r med n elementi. To pomeni, da je

$$C_n^{r\,(p)} = C_{n+r-1}^r = \left(\begin{array}{c} n+r-1 \\ r \end{array} \right).$$

2. Izračunajmo, kolikšna je verjetnost, da na lotu zadenemo sedmico. Izbrati moramo od 8 do 17 števil med številkami od 1 do 39. Recimo, da se odločimo, da bomo izbrali 17 števil. Vseh možnih kombinacij za izbiro sedmih števil na listku je

$$C_{39}^7 = \left(\begin{array}{c} 39 \\ 7 \end{array} \right) = 15380937.$$

Če izbiramo 17 številk, je različnih kombinacij sedmih števil lahko

$$C_{17}^7 = \left(\begin{array}{c} 17 \\ 7 \end{array} \right) = 19448.$$

Verjetnost, da bomo zadeli sedmico je

$$\frac{C_{17}^7}{C_{39}^7} = 0.001264422.$$

Če pa izbiramo le 8 števil, je $C_8^7 = 8$ in verjetnost zadetka

$$\frac{C_8^7}{C_{39}^7} = 0.0000005201,$$

torej 1 v dveh miljonih.

3. Letalske družbe pogosto prodajo več kart, kot je sedežev v letalu, ker pričakujejo, da si bo nekaj potnikov tik pred zdajci premislilo ali pa bodo zadržani. Po drugi strani pa družbe ne želijo, da bi prepogosto kdo od potnikov z veljavno karto ostal brez sedeža. Recimo, da je verjetnost, da se potnik s kupljeno karto pojavi, 0.9. Koliko več kart lahko letalska družba proda za letalo s 500 sedeži, da bo verjetnost, da bo kdo ostal brez sedeža, manjša od 5%? Kolikšna je verjetnost, da kdo od potnikov ostane brez sedeža, če prodajo 550 kart?

4. Tovarna izdeluje nek izdelek. Poslovodja želi vedeti, kolikšen je delež slabih izdelkov. Postopek kontrole kvalitete izdelka je tak, da se pri tem izdelek uniči. Kako naj poslovodja izbere izdelke za kontrolo in koliko, da bo lahko z verjetnostjo 99% trdil, da je delež slabih izdelkov celotne tovarne enak deležu slabih izdelkov med izbranimi?

Za reševanje zadnjih dveh problemov potrebujemo nekoliko več znanja iz verjetnosti in statistike. Rešili ju bomo ob koncu poglavja.

2. Osnovni pojmi in računanje z dogodki

Osnovna objekta verjetnostnega računa sta **poskus** in **dogodek**. Poskus je npr. met kocke ali pa izbira kroglice izmed množice kroglic. Dogodek je pri metu kocke npr. to, da pade 4, pri izbiri kroglic pa to, da izberemo kroglico s številko 10. Dogodke bomo označevali z velikimi črkami z začetka abecede, $A, B, C, A_1, A_2...$, poskuse pa z velikimi črkami s konca abecede, $X, Y, Z, X_1, X_2, ...$ Označimo z A_i dogodek, da pri metu kocke pade število i, z G dogodek, da pri metu kocke pade število med 1 in 6, N naj bo dogodek, da pri metu ne pade nobeno od števil od 1 do 6 in B dogodek, da pri metu kocke pade sodo število. Dogodek G se zgodi pri vsakem metu kocke, dogodek N nikoli, dogodki A_i in B pa včasih.

Definicija 2.3. *Dogodek, ki se zgodi v vsaki ponovitvi poskusa X se imenuje* **gotov dogodek**. *Označimo ga z G. Dogodek, ki se ne zgodi v nobeni ponovitvi poskusa, se imenuje* **nemogoč** *in ga označimo z N. Dogodku, ki se v kaki ponovitni poskusa zgodi, v kaki pa ne, pravimo* **slučajen**.

Naj se dogodek A v n ponovitvah poskusa zgodi k-krat. Število k je **frekvenca** *dokodka A v n ponovitvah poskusa X.* **Relativna frekvenca** *dogodka A v n poskusih, $f_n(A)$, je razmerje med frekvnco in številom poskusov,*

$$f_n(A) = \frac{k}{n}.$$

Verjetnost $P(A)$ *dogodka A je* $\lim_{n\to\infty} f_n(A)$.

Posledica 2.4. *Za vsak dogodek A je $0 \leq P(A) \leq 1$. Gotov dogodek ima verjetnost* 1, *nemogoč pa* 0.

Posledica 2.5. *Naj ima poskus X s možnih izidov, ki so enako verjetni. Če je za dogodek A ugodnih k izidov, je $P(A) = k/s$.*

Primeri.
1. Izračunaj verjetnost, da pri metu poštene kocke pade 3, verjetnost, da pade sodo število in verjetnost, da pade število, ki je deljivo s 3. Označimo prvi dogodek z A_3, drugega z B in tretjega s C.

$$P(A_3) = \frac{1}{6}, \;\; P(B) = \frac{3}{6}, \;\; B(C) = \frac{2}{6}.$$

2. Imamo nepošten kovanec. Verjetnost, da bo padel grb je 1/4, verjetnost, da pade cifra, pa 3/4. Kolikšna je verjetnost, da bomo v treh zaporednih metih vrgli enkrat cifro in dvakrat grb? Za to imamo naslednje možnosti: 1. najprej vržemo cifro in dvakrat grb, 2. najprej vržemo grb, potem cifro in še enkrat grb, 3. najprej vržemo dvakrat grb in potem cifro. Vsaka od teh kombinacij je enako verjetna in ima verjetnost

$$p = \frac{1}{4}\frac{1}{4}\frac{3}{4}.$$

Verjetnost $P(A)$ je enaka $3p$.

Pri poskusu X se zgodijo dogodki $A, B, C, ...$, ki imajo verjetnosti $P(A)$, $P(B)$, itd. Kako bi iz tek verjetnosti izračunali, kakšna je verjetnost dogodka, ki je v zvezi z danimi dogodki? Oglejmo si ta problem na konkretniem primeru.

Dane so naslednje množice:

$$\begin{aligned}
\mathcal{G} &= \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}, \\
\mathcal{A} &= \{2, 4, 6, 8, 10\}, \\
\mathcal{B} &= \{1, 3, 4, 8\}, \\
\mathcal{C} &= \{5, 7, 9\} \text{ in} \\
\mathcal{D} &= \{4, 8\}.
\end{aligned}$$

Iz množice $\mathcal{G}$ na slepo izberimo število x. Dogodek A ze zgodi, če je $x \in \mathcal{A}$. B je dogodek, da je $x \in \mathcal{B}$, C je dogodek, da je $x \in \mathcal{C}$, D je dogodek, da je $x \in \mathcal{D}$ in G je dogodek, da je $x \in \mathcal{G}$. Z N označimo nemogoč dogodek. Zapišimo verjetnosti: $P(A) = 4/10$, $P(B) = 5/10$, $P(C) = 3/10$, $P(D) = 2/10$, $P(G) = 1$.

(1) Dogodek D zgodi vedno, ko se zgodi A; pravimo, da je D **način** dogodka A in to zapišemo tako: $D \subset A$. Če je D način dogodka A, je $P(D) \leq P(A)$. V našem primeru je $P(D) = 2/10$, $P(A)$ pa je $4/10$.

(2) Če se dogodka E in F vedno zgodita hkrati, pravimo, da sta **enaka**, $E = F$. Imamo bele lesene krogle in črne železne krogle. Naj bo E dogodek, da smo izbrali belo kroglo, F pa dogodek, da smo izbrali leseno. Dogodka E in F se vedno zgodita hkrati, zato sta enaka. Enaka dogodka imata enaki verjetnosti.

(3) Dogodek, da se zgodita A in B hkrati, imenujemo **produkt** dogodkov in ga označimo z AB ali $A \cap B$. V našem primeru je $AB = D$. Izračunajmo še verjetnost: $P(D) = 2/10$. Pozor: $P(AB)$ je v splošnem različen od $P(A)P(B)$.

(4) Naj bo E dogodek, da se od dogodkov A in B zgodi vsaj eden. To pomeni, da se dogodek E zgodi, če je izbrano število v množici $\mathcal{A} \cup \mathcal{B}$. Dogodek E imenujemo **vsota** dogodkov A in B. Uporabljamo oznaki $C = A + B$ ali $C = A \cup B$. Velja: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. V našem primeru je $P(AB) = 4/10 + 5/10 - 2/10 = 7/10$.

(5) Dogodka A in C se ne moreta zgoditi hkrati. Taka dva dogodka imenujemo **nezdružljiva**. Njun produkt je torej nemogoč dogodek, $AC = N$.

(6) Naj bo $E = A+B$. Dogodka E in C se ne moreta zgoditi hkrati; pri vsakem poskusu ali se zgodi ali eden ali drugi. Velja: $EC = N$ in $E+C = G$. Dogodek E imenujemo **negacija** dogodka C, $\overline{C} = E$. Prav tako je tudi C negacija E, $\overline{E} = C$. Izračunajmo še verjetnosti: $P(E + C) = P(E) + P(C) - P(EC) = 1$. Ker je EC nemogoč dogodek, je $P(C) = 1 - P(E)$.

(7) Če je verjetnost produkta dogodkov enaka produktu verjetnosti dogodkov, pravimo, da sta dogodka **neodvisna**. Dogodka A in B sta neodvisna, saj je $P(AB) = 2/10 = 4/10 \cdot 5/10$.

(8) Naj bo A_i dogodek, da smo izbrali število i, $1 \leq i \leq 10$. Očitno je za vsak $i \neq j$ produkt A_iA_j nemogoč dogodek. Vsak dogodek A, B, C, D lahko izrazimo kot vsoto teh dogodkov. Pravimo, da so ti dogodki **sestavljeni**. Dogodkov A_i pa ne moremo zapisati kot (netrivialno) vsoto dogodkov, tj. kot vsoto dogodkov, od katerih nobeden ni dogodek N. Pravimo, da so dogodki A_i **elementarni**. V vsakem poskusu se zgodi natanko en dogodek izmed $A_1, \ldots, A_{10}$. Pravimo, da sestavljajo ti dogodki **popoln sistem dogodkov**. Velja: $A_1 + \ldots + A_{10} = G$ in $P(A_1) + \ldots + P(A_{10}) = 1$.

Primer. **Paradoks Chevaliera de Mereja.** Imamo naslednji dve igri: pri prvi vržemo kocko štirikrat in zmagamo, če je bila vsaj ena šestica, pri drugi pa vržemo dve kocki štiriindvajsetkrat in zmagamo, če smo vsaj enkrat hkrati vrgli dve šestici. De Mere je izračunal, da je verjetnost zmage pri prvi igri $4/6$, pri drugi pa $24/36 = 4/6$. Razmišljal je takole: recimo, da smo v prvem metu vrgli 6, v naslednjih pa karkoli. Verjetnost tega dogodka je $1/6$. Naslednja

možnost je, da smo pri prvem metu vrgli karkoli in 6 pri drugem, kar ima tudi verjetnost 1/6 itd. Ker imamo 4 take kombinacije, je verjetnost, da je padla vsaj ena šestica, enaka 4/6. Podobno je razmišljal za drugi primer. Vseh možnih izidov metov dveh kock je 36. To pomeni, da ima vsak izid verjetnost 1/36. Recimo, da smo prvič vrgli dve šestici, v naslednjih metih pa karkoli. Verjetnost tega dogodka je 1/36. Recimo, da smo v prvem metu vrgli karkoli, v drugem pa dve šestici. Verjetnost tega dogodka je spet 1/36 itd. Skupna verjetnost moora biti 24/36.

Kockarska praksa je pokazala, da igri nista enakovredni. Kje je napaka? Če natančneje pogledamo, kako ke De Mere štel ugodne izide, vidimo, da je nekatere štel po večkrat. Pravilno razmišljanje za prvi primer bi bilo takole: v prvem metu vržem 6 v ostalih karkoli, verjetnost je 1/6. Drugi korak pa je malo drugačen. Če rečemo, da smo v prvem metu vrgli karkoli, v drugem pa 6 (in v zadnjih dveh karkoli), smo še enkrat šteli možnost, da smo v prvem metu vrgli šest. Pravilno bi bilo razmišljati tako: v prvem metu nismo vrgli šestice, v drugem smo vrgli šestico, v zadnjih dveh pa karkoli. Verjetnost tega dogodka je $5/6 \cdot 1/6$. Za naslednji primer ne prvič ne drugič ne vržemo šestice, vržemo jo tretjič, četrtič pa pade karkoli. Verjetnost tega dogodka je $5/6 \cdot 5/6 \cdot 1/6$. Verjetnost dogodka, da vržemo šestico šele v zadnjem metu, je $5/6 \cdot 5/6 \cdot 5/6 \cdot 1/6$. Verjetnost, da zmagamo, je vsota teh štirih verjetosti. Lahko razmišljamo tudi drugače. Nasprotni dogodek tega, da pade vsaj ena šestica je, da ni padla nobena šestica. Verjetnost tega dogodka je $(5/6)^4$, zato je verjetnost dogodka, da pade vsaj ena šestica enaka $1 - (5/6)^4 = 0,5177$.

Podobno je za drugo igro. Verjetnost, da nikoli ne padeta dve šestici, je $(35/36)^{24}$. Verjetnost, da sta dve šestici padli vsaj enkrat, je $1 - (35/36)^{24} = 0,4914$.

3. Pogojna verjetnost

Spet si poglejmo igro s kockami. Vržemo dve kocki. Kolikšna je verjetnost, da smo vrgli šestico, če vemo, da je vsota pik enaka 9? Naj bo A dogodek, da je padla šestica, B pa dogodek, da je vsota pik enaka 9. Izračunali bi radi verjetnost, da se je zgodil A pri pogoju B.

Definicija 3.2. Pogojna verjetnost *dogodka A pri pogoju B je verjetnost, da se zgodi A, če vemo, da se je zgodil B. Za pogojno verjetnost uporabljamo oznako $P(A/B)$.*

Preštejmo vse možne izide, pri katerih je vsota 9 : $(3,6)$, $(4,5)$, $(5,4)$ in $(6,3)$. V dveh od teh je padla šestica, zato je verjetnost $P(A/B) = 1/2$. Izračunali smo, kolikšen je delež množice A v množici B, torej razmerje med $P(A \cap B)$ in $P(B)$. Od tod sledi formula za izračun pogojne verjetnosti:

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

Če sta dogodka A in B neodvisna, je $P(A/B) = P(A)$. Iz zgornje formule dobimo še eno:

$$P(A/B)P(B) = P(AB) = P(B/A)P(A).$$

Če jo še malce premečmo, pa naslednjo:

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

Primeri.
1. Kolikšna je verjetnost, da je padla šestica pri vsotah $7, 8, 9, 10, 11, 12$? Označimo te dogodke $B_7, \ldots B_{12}$, dogodek, da pade šestica pa z A. Za vsoto 7 imamo 6 možnosti in šestica se pojavi v dveh, zato je $P(A/B_7) = 1/3$. Za vsoto 8 je 5 možnosti: $P(A/B_8) = 2/5$, za vsoto 10 so 3 možnosti in pri dveh

pade šestica, $P(A/B_{10}) = 2/3$, za vsoto 11 sta dve možnosti in pri obeh je padla šestica, $P(A/B_{11}) = 1 = P(A/B_{12})$.

2. Zdaj pa obrnimo vprašanje. Recimo, da vemo, da je padla šestica. Kolikšna je verjetnost, da je vsota enaka 9?

3. Naj bo A dogodek, da vržemo števili 3 ali 6, B pa dogodek, da je padlo sodo število. Kolikšna je verjetnost, da se je zgodil A, če vemo, da se je zgodil B? Ker vemo, da se je zgodil B, je število pik sodo, torej imamo na voljo $2, 4, 6$. Od teh izidov je za dogodek A ugoden samo eden, in sicer 6. Zato je pogojna verjetnost $P(A/B) = 1/3$.

4. Slučajne spremenljivke, matematično upanje, standardni odklon

Za krajše zapisovanje verjetnosti izidov pri poskusu X si pomagamo z **verjetnostnimi shemami**:

$$X : \begin{pmatrix} A_1 & A_2 & \dots & A_k \\ p_1 & p_2 & \dots & p_k \end{pmatrix}.$$

Ta shema pove, da je verjetnost, da se v poskusu X zgodi dogodek A_i, enaka p_i :

$$P(X = A_i) = p_i.$$

Vsota vseh p_i mora biti 1 in $p_i \geq 0$ za vsak i.

Definicija 4.2. *Poskusom X, katerih izidi so števila, pravimo* **slučajne spremenljivke**.

Primeri.
1. Zapiši verjetnostno shemo za met kocke.

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

Zapiši verjetnostno shemo za slučajno spremenljivko $Y = (X-3)^2$. Najprej si oglejmo, kakšne možne izide ima spremenljivka Y : če je $X = 1$ ali $X = 5$ je $Y = 4$, če je $X = 2$ ali 4 je $Y = 1$, če je $X = 3$ je $Y = 0$ in če je $X = 6$ je $Y = 9$. Shema je

$$Y : \begin{pmatrix} 0 & 1 & 4 & 9 \\ 1/6 & 2/6 & 2/6 & 1/6 \end{pmatrix}.$$

2. Zapiši verjetnostno shemo za met kovanca, kjer je verjetnost, da pade cifra enaka 3/4. Označimo z 1 izid, da pade cifra in z 0 izid, da pade grb.

$$X : \begin{pmatrix} 0 & 1 \\ 1/4 & 3/4 \end{pmatrix}.$$

Vzemimo dva enaka kovanca, ki jima pripadata slučajni spremenljivki X_1, X_2 z enako porazdelitvijo kot zgoraj. Napiši shemo za spremenljivko $Y = X_1X_2$!

$$Y : \begin{pmatrix} 0 & 1 \\ 7/16 & 9/16 \end{pmatrix}.$$

Zapiši še shemo za slučajno spremenljivko $Z = X_1 + X_2 - 1$.

$$Z : \begin{pmatrix} -1 & 0 & 1 \\ 1/16 & 6/16 & 9/16 \end{pmatrix}.$$

Definicija 4.3. *Naj bo X poskus, kjer izbiramo števila od* 1 *do n z verjetnostno shemo*

$$X : \begin{pmatrix} 1 & 2 & \dots & n \\ p_1 & p_2 & \dots & p_n \end{pmatrix};$$

A_i pomeni dogodek, da smo izbrali število i. Izraz

$$E(X) := \sum_1^n ip_i$$

imenujemo **pričakovana vrednost** *ali* **matematično upanje**. *Izraz*

$$D(X) := E((X - E(X))^2) = E(X^2) - E(X)^2$$

imenujemo **varianca**, *izraz*

$$\sigma(X) = \sqrt{D(X)}$$

pa **standardni odklon**.

Opomba. Matematično upanje nam pove "povprečno vrednostšlučajne spremenljivke, varianca pa povprečje kvadratov razlik med matematičnim upanjem in našo spremenljivko; lahko bi rekli, da meri "razpršenost podatkov"glede na povprečje.

Primeri. Izračunaj E in σ za slučajne spremenljivke iz zgornjih primerov!

1. Met poštene kocke.

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

$$\begin{aligned} E(X) &= \frac{1}{6}(1+2+3+4+5+6) = 3.5, \\ D(X) &= E(X^2) - E(X)^2 = \frac{1}{6}(1+4+9+16+25+36) - \frac{49}{4} \\ &= \frac{91}{6} - \frac{49}{4} = 2.92, \\ \sigma(X) &= 1.71. \end{aligned}$$

2.

$$Y : \begin{pmatrix} 0 & 1 & 4 & 9 \\ 1/6 & 2/6 & 2/6 & 1/6 \end{pmatrix}.$$

$$\begin{aligned} E(Y) &= 2/6 + 8/6 + 9/6 = 19/6, \\ D(Y) &= E(X^2) - E(X)^2 = 2/6 + 32/6 + 81/6 - 19^2/36 = 9.14, \\ \sigma(Y) &= 3.02. \end{aligned}$$

3.

$$Z : \begin{pmatrix} -1 & 0 & 1 \\ 1/16 & 6/16 & 9/16 \end{pmatrix}.$$

$$\begin{aligned} E(Z) &= -1/16 + 9/16 = 1/2, \\ D(Z) &= E(Z^2) - E(Z)^2 = 5/8 - 1/4 = 3/8, \\ \sigma(Z) &= 0.61. \end{aligned}$$

Dva poskusa sta med sabo **neodvisna**, če je vsak dogodek prvega poskusa neodvisen od kateregakoli dogodka iz drugega poskusa. Naj ima poskus X izide A_i, poskus Y pa izide B_j. Po definiciji sta poskusa neodvisna, če je

$$P(X = A_i, Y = B_j) = P(X = A_i)P(Y = B_j) \text{ za vsak par } i, j.$$

Poskusoma, ki nista neodvisna, pravimo **odvisna**. Enaka definicija je za neodvisnost slučajnih spremenljivk, saj so slučajne spremenljivke le poskusi s številskimi izidi.

Primeri.
1. Met dveh kock. Očitno imamo opravka z dvema neodvisnima poskusoma. Naj bo X met prve kocke, Y pa met druge kocke. Verjetnost dogodka A_i, da smo s prvo kocno vrgli število i je $1/6$, verjetnost B_j dogodka, da smo z drugo kocko vrgli j je $1/6$ ($1 \leq i, j \leq 6$). Verjetnost dogodka $P(X = A_i, Y = B_j)$ je $1/36$.

2. Spet mečemo dve kocki, le da sta med seboj povezani z vrvico. Z X označimo met prve kocke z Y met druge in z A_i, B_j iste dogodke kot zgoraj. Verjetnosti dogodkov $P(X = A_i, Y = B_j)$ popisuje spodnja tabela.

X \ Y	1	2	3	4	5	6
1	2/72	1/72	0	3/72	1/72	3/72
2	1/72	4/72	1/72	2/72	1/72	3/72
3	1/72	1/72	2/72	2/72	3/72	2/72
4	3/72	2/72	1/72	4/72	3/72	2/72
5	1/72	1/72	2/72	1/72	4/72	5/72
6	1/72	3/72	4/72	0	1/72	1/72

Izračunaj verjetnost, da pade na prvi kocki 4, pri pogoju, da je na drugi kocki padla 2. Verjetnost, da je na drugi kocki padla 2 je $12/72 = 1/6$. Verjetnost dogodka $P(X = 4, Y = 2)$ je $2/72 = 1/36$. Verjetnost

$$P(X = 4/Y = 2) = \frac{1/36}{1/6} = \frac{1}{6}.$$

Verjetnost, da je na drugi kocki padla 3 pri pogoju, da je na prvi kocki padla 2 je

$$P(Y = 3/X = 2) = \frac{1/72}{12/72} = \frac{1}{12}.$$

Naj bo $Z = X + Y$. Verjetnost, da je padla vsota 5 je $P(Z = 5) = 8/72 = 1/9$. Izračunaj $E(Z), \sigma(Z)$.

$$Z : \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{2}{72} & \frac{2}{72} & \frac{5}{72} & \frac{8}{72} & \frac{8}{72} & \frac{9}{72} & \frac{15}{72} & \frac{10}{72} & \frac{6}{72} & \frac{6}{72} & \frac{1}{72} \end{pmatrix}.$$

5. Zaporedja neodvisnih dogodkov

Oglejmo si naslednjo igro z ruleto: stavimo en žeton na rdeče. Vseh polj na ruleti je 37, od teh je 18 rdečih, 18 črnih in eno zeleno. Če pade krogla na zeleno polje, žeton pobere igralnica. Če pade kroglica na rdeče polje, nam igralnica da žeton nazaj in še en žeton zraven, tako da pridobimo en žeton. V nasprotnem primeru žeton izgubimo. Verjetnost, da igro dobimo je $p = 18/37$, verjetnost, da izgubimo, pa je $q = 19/37$. Verjetnostna shema, ki popisuje igro je

$$X : \begin{pmatrix} -1 & 1 \\ 19/37 & 18/37 \end{pmatrix}.$$

Kolikšna je verjetnost, da bomo po $n = 100$ igrah imeli dobiček? Da bi ta problem rešili, najprej izračunajmo verjetnost dogodka, da v $n = 100$ igrah k-krat zmagamo:

$$P_n(k) = \binom{n}{k} p^k q^{n-k}.$$

Koliko žetonov imamo, če smo k-krat zmagali? Dobili smo jih k izgubili pa $n - k$, torej je skupna vsota $k - n + k = 2k - n$. Označimo z X_n slučajno spremenljivko, ki popisuje n iger na ruleti. Izračunali smo

$$P(X_n = 2k - n) = P_n(k).$$

Če označimo i-to igro na ruleti z X^i, je

$$X_n = X^1 + X^2 + \ldots + X^n;$$

če $k-$krat zmagamo v igrah $X^1, \ldots, X^n$, to pomeni, da je v vsoti k enic,

$$X_n = k - (n - k) = 2k - n.$$

Vsaka od spremenljivk X^i je porazdeljena enako, kot spremenljivka X.

Narišimo izide in pripadajoče verjetnosti v histogram. Ploščine stebričkov nad številom $2k - n$ predstavljajo verjetnost zaslužka $2k - n$, torej $P_n(k)$.

Vsota ploščin vseh stebričkov je enaka 1. Zanima nas verjetnost, da bo naš dobiček večji od 0, torej ploščina desno od ničle.

Vidimo, da je ploščina dela histograma skoraj ploščina pod grafom zvončaste krivulje. Izkaže se, da gre za krivuljo

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Krivulja se imenuje **normalna krivulja** s parametroma μ in σ. Oznaki $\sigma = \sigma(X_n)$ in $\mu = E(X_n)$ pomenita standardni odklon in matematično upanje. Krivulja ima vrh pri $x = E(X_n)$. Večji, kot je σ, nižji in bolj širok je zvonec. Verjetnost $P_n(k)$ je približno enaka

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(k-\mu)^2}{2\sigma^2}}.$$

Posebno pozornost si zasluži krivulja za parametra $\mu = 0$ in $\sigma = 1$:

$$f(x;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Ploščina pod krivuljo nad intervalom $[-1,1]$ je 68%, ploščina nad intervalom $[-2,2]$ je 95.45%, ploščina nad intervalom $[-3,3]$ pa je 99.73%. Ploščine so tabelirane v matematičnem priročniku.

Izrek 5.2. *Dana je normalna krivulja $f(x;\mu,\sigma)$. Nad intervalom $[\mu-\sigma,\mu+\sigma]$ je 68% ploščine, nad intervalom $[\mu-2\sigma,\mu+2\sigma]$ je 95.45% ploščine, nad intervalom $[\mu-3\sigma,\mu+3\sigma]$ pa 99.73% ploščine. Ploščina pod krivuljo $f(x;\mu,\sigma)$ nad intervalom $[a,b]$ je enaka ploščini pod standardno normalno krivuljo nad intervalom $[(a-\mu)/\sigma,(b-\mu)/\sigma]$.*

Dokaz. Izrek o substituciji. V integral

$$I = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

uvedemo novo spremenljivko

$$u = \frac{(x-\mu)}{\sigma}; \quad du = \frac{dx}{\sigma}.$$

Dobimo

$$I = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} du.$$

$\diamondsuit$

Izračunajmo še ploščino, ki nas zanima. Potrebujemo matematično upanje in standardni odklon. Preden pa se lotimo nadaljnega računanja, zapišimo naslednje vsote:

$$\begin{aligned}\sum_0^n P_n(k) &= 1,\\ \sum_0^n kP_n(k) &= np,\\ \sum_0^n k^2P_n(k) &= n(n-1)p^2+np.\end{aligned}$$

$$\begin{aligned}E(X_n) &= \sum_0^n(2k-n)P_n(k)\\ &= 2\sum_0^n kP_n(k) - n\sum_0^n P_n(k)\\ &= 2np-n,\\ D(X_n) &= \sum_0^n(2k-n)^2P_n(k) - (2np-n)^2\\ &= 4\sum_0^n k^2P_n(k) - 4n\sum_0^n kP_n(k) + n^2\sum_0^n P_n(k) - (4n^2p^2+n^2-4n^2p)\\ &= 4n(n-1)p^2+4np-4nnp+n^2-4n^2p^2-n^2+4n^2p\\ &= -4np^2+4np = 4npq,\\ \sigma(X_n) &= 2\sqrt{npq}.\end{aligned}$$

Vstavimo $n = 100$ in p, q :

$$\begin{aligned}E(X_{100}) &= 100(\frac{36}{37}-1) = -100\frac{1}{37} = -2.70,\\ \sigma(X_{100}) &= 2\cdot 10\cdot 0.4998 = 9.996.\end{aligned}$$

Standardni odklon zaokrožimo na 10. Zanima nas ploščina pod normalno krivuljo desno od

$$x = \frac{0-E(X_n)}{\sigma(X_n)} = \frac{2.7}{10} = 0.27.$$

Iz tabele preberemo:

$$P = 0,393 = 39.3\%.$$

Verjetnost, da bomo po 100 igrah kaj dobili, je 39.9%.

Zapišimo še formule za standardni odklon in matematično upanje, če je začetna porazdelitev

$$X : \begin{pmatrix} a & b \\ p & q \end{pmatrix}.$$

Dobimo

$$E(X_n) = n(ap + bq) = nE(X), \qquad \sigma(X_n) = |a - b|\sqrt{npq} = \sqrt{n}\sigma(X).$$

Primeri.

1. Letalske družbe pogosto prodajo več kart, kot je sedežev v letalu, ker pričakujejo, da si bo nekaj potnikov tik pred zdajci premislilo ali pa bodo zadržani. Po drugi strani pa družbe ne želijo, da bi prepogosto kdo od potnikov z veljavno karto ostal brez sedeža. Recimo, da je verjetnost, da se potnik s kupljeno karto pojavi, 0.9. Koliko več kart lahko letalska družba proda za letalo s 500 sedeži, da bo verjetnost, da bo kdo ostal brez sedeža, manjša od 5%? Kolikšna je verjetnost, da kdo od potnikov ostane brez sedeža, če prodajo 550 kart?
Začetna spremenljivka X ima porazdelitev

$$X : \begin{pmatrix} 0 & 1 \\ 0.1 & 0.9 \end{pmatrix},$$

Po zgornjih formulah je

$$E(X_n) = n \cdot 0.9, \;\; \sigma(X_n) = \sqrt{n}\sqrt{0.09} = \sqrt{n}0.3.$$

Spremenljivno standardiziramo:

$$Y_n = \frac{X_n - E(X_n)}{\sigma(X_n)}.$$

Odgovorimo najprej na drugo vprašanje. Prodali smo $n = 550$ kart. V tem primeru je

$$E(X_n) = 550 \cdot 0.9 = 495, \;\; \sigma(X_n) = 7.03$$

Kolikšna je verjetnost, da bo $X_n \leq 500$? Računamo ploščino

$$\int_{-\infty}^{500} f(x; E(X_n), \sigma(X_n))dx.$$

V standardnih enotah

$$\begin{aligned} \frac{500 - 495}{7.03} &= 0.71, \\ \int_{-\infty}^{0.71} f(x; 0, 1) &= 76\%. \end{aligned}$$

Verjetnost, da bo prišlo največ 500 potnikov je 76%. V 24% primerov bo prišlo preveč potnikov.

Zanima nas, pri katerem n je verjetnost dogodka $X_n \leq 500$, vsaj 95%. Ugotoviti moramo, pri katerem n je ploščina pod grafom

$$\int_{-\infty}^{500} f(x; E(X_n), \sigma(X_n))dx \geq 95\%.$$

Po prejšnjem je to enako

$$\int_{-\infty}^{a} f(x; 0, 1)dx \geq 95\%$$

za

$$a = \frac{500 - E(X_n)}{\sigma(X_n)}.$$

Pogledamo v tabele in preberemo, da je

$$a = 1.65.$$

Izračunajmo, kakšen mora biti n :

$$1.65 = \frac{500 - n \cdot 0.9}{\sqrt{n} \cdot 0.3}.$$

Pišimo $u = \sqrt{n}$:

$$u^2 0.9 + 0.3 \cdot 1.65u - 500 = 0.$$

Edina rešitev kvadratne enačbe, ki pride v poštev, je

$$u = \frac{-0.495 + \sqrt{0.24 + 1800}}{1.80} = 23.2968,$$

saj je druga negativna. Kvadriramo u in dobimo pravo število:

$$n = 543.$$

2. Zavarovalnica ima 10000 zavarovancev. Znano je, da je povprečje zahtevkov 2.4 enote, standardni odklon pa 1.6 enote (enota je 100.000 SIT). Kolikšna je verjetnost, da bo skupna višina zahtevkov pri 1000 zahtevkih presegala 2500 enot?

V zavarovalnici pričakujejo letno 2000 zahtevkov. Kolikšna mora biti zavarovalna premija, če želimo, da je verjetnost, da bo skupna vsota zahtevkov večja od skupne vsote premij, manjša od 0.5%?

Odgovor na prvo vprašanje da integral

$$\int_{-\infty}^{2500} f(x; E(X_{1000}), \sigma(X_{1000}))\, dx$$

Pretvorimo ga v standardne enote:

$$\begin{aligned} E(X_{1000}) &= 1000 \cdot 2.4 = 2400, \\ \sigma(X_{1000}) &= 1.6\sqrt{1000} = 50.6, \\ \frac{2500 - 2400}{50.6} &= 1.97. \end{aligned}$$

Dobimo

$$\int_{-\infty}^{1.97} f(x; 0, 1)\, dx = 97.6\%.$$

Verjetnost, da bo skupna višina zahtevkov presegla 2500 enot, je 2.4%. Kolikšna je verjetnost, da bo (pri istem številu zahtevkov) skupna vsota presegla 2800 enot?

Rešimo še drugi del naloge. Najprej moramo ugotoviti, za kateri a je verjetnost $P(X_{2000} < a) > 99.5\%$. Količina a bo skupna vsota vseh zahtevkov. Pretvorimo vse v standardne enote:

$$\begin{aligned} E(X_{2000}) &= 2000 \cdot 2.4 = 4800, \\ \sigma(X_{2000}) &= 1.6\sqrt{2000} = 71.5, \\ \frac{a - 4800}{71.5} &= b. \end{aligned}$$

Dobimo

$$\int_{-\infty}^{b} f(x; 0, 1)\, dx = 99.5\%.$$

Iz tabel preberemo, da je $b = 2.57$. Izračunamo še a :

$$a = 4983.9.$$

Premija mora biti 0.4983 enote ali 49839 SIT.

6. Enostavno slučajno vzorčenje

V uvodu tega poglavja smo se srečali s problemom določanja deleža slabih izdelkov v proizvodnji. Podoben problem temu so javnomenjske ankete, ocenjevanje rasti življenskih stroškov, revizorsko vzorčenje vknjižb itd. Ocene, ki jih dobimo na podlagi izbranega vzorca, so seveda le približek dejanskega stanja, zato se moramo vprašati, kako natančne so naše ocene. V tem razdelku se bomo ukvarjali z vprašanjem, kako je potrebno izbrati vzorec in kaj lahko potem trdimo o natančnosti ocen.

1. **Plebiscit** 1990. Na FDV so pred plebiscitom izpeljali anketo SJM90 (Slovensko javno mnenje 1990), v kateri so med drugim ugotavljali, kakšno je razpoloženje Slovencev do odcepitve (O) in samostojnosti (S). Vzorec je zajel 2074 volilcev. Rezultati ankete so naslednji:

O \ S	da	ne	ostalo
da	1306	11	34
ne	183	125	63
ostalo	110	12	230

Prvi korak pri razmišljanju o zanesljivosti ocen na podlagi vzorca je opis načina izbire vzorca, **vzorčni načrt**. Pri ustrezno izpeljanih anketah je način vzorčenja natančno predpisan. Pri anketah SJM je osnovni okvir vzorčenja centralni register prebivalstva pri Statističnem uradu RS, iz katerega dobimo spisek volilcev. Ta je za potrebe vzorčenja razdeljen po geografskem ključu na dele po 4200 volilcev. Te skupine so primarne vzorčne enote. Vsaka od primarnih vzorčnih enot je spet razdeljena na manjše, sekundarne vzorčne enote po 100 volilcev. Vzorčenje pri SJM poteka tako: na prvem koraku anketarji izberejo 140 primarnih vzorčnih enot, v drugem koraku iz vsake izbrane primarne enote naključno izberejo še 3 sekundarne enote, nazadnje pa iz vsake sekundarne enote naključno izberejo 5 volilcev. Skupno je v vzorec izbranih 2100 volilcev. Če anketarjem v petih poskusih ne uspe dobiti odgovora od izbranega volilca, neznani odgovor obravnavajo kot manjkajoč

podatek. Videli bomo, da tak način vzorčenja zagotavlja visoko natančnost naših ocen.

2. Predsedniške volitve v ZDA 1936. Oglejmo si še primer ankete, pri kateri so bili rezultati zelo narobe. Pred predsedniškimi volitvami leta 1936 je prestižna revija Literary Digest izvedla javnomnenjsko anketo in napovedala, da bo na volitvah zmagal A. Landon, F.D. Roosevelt pa izgubil. Naslednja tabela prikazuje napovedi revije in dejanske rezultate.

	napoved LD	dejanski rezultat
A. Landon (R)	57 %	38 %
F.D. Roosevelt (D)	43 %	62 %

Ko pogledamo, kaj je bilo z vzorčenjem, je jasno, zakaj so bili izidi tako zelo narobe. Izvor za izbiro volilcev so bili telefonski imeniki, spiski članov elitnih klubov in podobno. To v letu 1936, ko je bila gospodarska kriza najhujša in je brezposelnost presegla 11 miljonov (25% delovne sile), gotovo ni bil reprezentativen vzorec. Poleg tega so vprašalnike poslali izbranim po pošti; od 10 miljonov poslanih anket so dobili le 2.4 miljona odgovorov, kar je znamenje za alarm. Če upoštevamo, da republikansko stranko podpirajo praviloma bogatejši sloji, je izid ankete kaj lahko napovedati. Anketa je bila zastavljena tako, da je v vzorec izbrala predvsem tiste, ki so bili v času krize kar precej bogati. Velikost vzorca tu nič ne pomaga, saj se le ponavlja ena in ista napaka. Revija je kmalu za tem propadla.

Omenimo še, da je v tistem času mlad statistik George Gallup na podlagi vzorca 5000 volilcev napovedal zmago FDR s 56 procenti; na podlagi vzorca 3000 anket revije LD je tudi napovedal izid ankete, in sicer, da bo FDR dobil 44% glasov.

Pri **enostavnem slučajnem vzorčenju** je verjetnost, da je enota zajeta v vzorec, enaka za vse enote. Enako verjetni so tudi vsi možni vzorci n enot populacije z N enotami.

Slovenija ima okoli 1500000 volilcev. Koliko je možnih vzorcev po 1000 volilcev? Po binomski formuli je to

$$\binom{1500000}{1000} \approx 2.19 \cdot 10^{3608}.$$

Oglejmo si konkreten primer. Izvajamo predvolilno anketo za lokalnega župana v populaciji velikosti $N = 1000000$ in izbiramo vzorec velikosti $n =$

1000. Videli smo, da je možnih vzorcev veliko. Zamislimo si virtualnega anketarja, ki izbere vzorec in izračuna, koliko procentov volilcev podpira župana. Virtualnemu anketarju zdaj naročimo, naj izdela ocene za vse možne vzorce in naj rezultate nariše v histogram. Histogramu se prilega zvončasta krivulja. Ta ima vrh pri pravem odstotku volilcev, ki podpirajo župana. Označimo ta procent s p. Mi v naprej seveda ne vemo, koliko je pravi procent in ne vemo, kolikšen je standardni odklon normalne krivulje. Vendar pa lahko standardni odklon ocenimo po naslednji formuli:

$$\sigma = \frac{\sqrt{p(1-p)}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

Kvadratni koren $\sqrt{(N-n)/(N-1)}$ imenujemo popravni faktor. Če je $n \leq 10\%N$, je popravni faktor skoraj 1 in ga lahko zanemarimo. Izraz $\sqrt{p(1-p)}$ ima maksimum pri $p = 0.5$; $\sqrt{0,5^2} = 0.5$. Za vzorce, ki so manjši od 10% populacije, dobimo naslednjo zgornjo mejo za standardni odklon:

$$\sigma \leq \frac{0.5}{\sqrt{n}}.$$

Pri naših podatkih je $\sigma \leq 0.0158 = 1.6\%$. Recimo, da je dejanski $p = 0.3$ (tega še ne vemo). Naučili smo se, da je nad območjem $[p - 2\sigma, p + 2\sigma]$ kar 95% ploščine pod normalno krivuljo. Recimo, da bo v vzorcu delež volilcev, ki podpirajo župana, enak $\hat{p}$. Potem je verjetnost, da bo $\hat{p}$ ležal na območju $[p - 2\sigma, p + 2\sigma]$ enaka 95%. Ali drugače: prava vrednost p je z verjetnostjo 95% v intervalu $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$. Z našo prvo oceno lahko povemo, da se $\hat{p}$ z verjetnostjo 95% razlikuje od pravega p za največ 3.2%.

Trditev 6.2. *Naj dani vzorec velikosti n predstavlja manj kot* 10% *populacije. Potem lahko z verjetnostjo* 95% *trdimo, da se ocena $\hat{p}$ iz vzorca razlikuje od prave vrednosti p za največ* 3.2%.

Zdaj pa res izvedemo anketo in izračunamo, da je $\hat{p} = 28\%$. Vemo, da smo se z verjetnostjo 95% zmotili za največ 3.2 procenta. Ali lahko oceno izboljšamo? Uporabimo $\hat{p}$ za izračun približka za σ in dobimo

$$\hat{\sigma} = \frac{\sqrt{0.28 \cdot 0.62}}{\sqrt{1000}} = 0.013 = 1.3\%.$$

To pomeni, da je napaka manjša, največ 2.6%.

Zdaj pa rešimo primer iz uvoda. Tovarna izdeluje nek izdelek. Poslovodja želi vedeti, kolikšen je delež slabih izdelkov. Postopek kontrole kvalitete izdelka je tak, da se pri tem izdelek uniči. Kako naj poslovodja izbere izdelke za kontrolo in koliko, da bo lahko z verjetnostjo 99% trdil, da se delež slabih izdelkov celotne tovarne razlikuje od deleža slabih izdelkov med izbranimi za največ 1%?

Ker želimo trditi, da bo naša ocena z verjetnostjo 99% natančna na npr. 1%, moramo vzeti interval $[p - 3\sigma, p + 3\sigma]$, saj vemo, da ta interval pokrije 99% ploščine. Ker želimo natančnost na 1%, mora biti $3\sigma \leq 1\%$. Vzemimo najenostavnejšo oceno za σ :

$$3\sigma \leq 3\frac{0.5}{\sqrt{n}} \leq 1\%.$$

Izračunajmo n :

$$n \geq \left(\frac{3 \cdot 0.5}{0.01}\right)^2 = 2.25 \cdot 10^4.$$

Vzeti bi bilo potrebno kar 22500 vzorcev. To je nekam veliko. Poskusimo si pomagati s kakšnim dodatnim podatkom. Recimo, da vemo, da je delež slabih izdelkov manjši od 10%. To da boljšo oceno za σ :

$$\sigma \leq \frac{0.3}{\sqrt{n}}.$$

Izračunamo n;

$$n \geq \left(\frac{3 \cdot 0.3}{0.01}\right)^2 = 0.81 \cdot 10^4 = 8100.$$

Kolikšen n bi dobili, če bi se zadovoljili z verjetnostjo 95% in natančnstjo 2%? Za 95% je potrebno vzeti interval $[p - 2\sigma, p + 2\sigma]$. Dobimo

$$n \geq \left(\frac{2 \cdot 0.3}{0.02}\right)^2 = 0.09 \cdot 10^4 = 900.$$

Recimo, da smo izmerili $\hat{p} = 0.05$ pri $n = 900$. Kolikšna je napaka?

$$\hat{\sigma} = \sqrt{\frac{0.05 \cdot 0.95}{900}} = 0.0072,$$

torej je naša napaka z verjetnostjo 95% največ 1.44%.