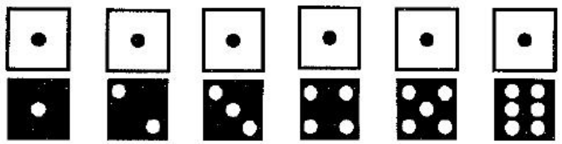
I.3. Pogojna verjetnost



Intriga (po Kvarkadabri)

Ne podcenjujmo vpliva najrazličnejših rubrik v popularnih časopisnih prilogah, kjer nas domnevni štrokovnjakižasipajo z nasveti vseh vrst,

- rubrike krojijo mnenja ljudi in spreminjajo navade celotnih nacij,

- sprožajo obsežne polemike tako med širšimi množicami kot tudi v ozki strokovni javnosti.

Na področju zdravja in prehrane tako burne odzive seveda pričakujemo, povsem nekaj drugega pa je, če jih sproži preprosto *matematično vprašanje*.

Revija *Parade* - kot prilogo jo vsako nedeljo dodajo več kot 400 ameriškim časopisom in doseže okoli 70 milijonov bralcev, že dolgo izhaja rubrika z imenom "Vprašajte Marilyn." Ureja jo **Marilyn vos Savant**.

Sredi 80ih jo je *Guinnessova knjiga rekordov* razglasila za rekorderko z najvišjim inteligenčnim količnikom na planetu.

V svoji rubriki zdaj že več kot 20 let odgovarja na najrazličnejša vprašanja bralcev in rešuje njihove težave.

Med vsemi vprašanji, ki jih je kdaj obravnavala, ima prav posebno mesto na prvi pogled zelo preprost problem, ki ji ga je 9. septembra 1990 zastavil gospod Craig F. Whitaker:

Dve kozi in avtomobil

"Vzemimo, da sodelujete v nagradni igri,
kjer vam ponudijo na izbiro troje vrat.

Za enimi se skriva avto, za drugima dvema pa koza.

Recimo, da izberete vrata številka 3,
voditelj igre, ki ve, kaj se nahaja za posameznimi vrati,
pa nato odpre vrata številka 1, za katerimi se pokaže koza.

Nato vas vpraša: 'Bi se sedaj raje odločili za vrata številka 2?'

Zanima me, ali se tekmovalcu splača zamenjati izbor vrat?"

Poudariti je potrebno, da mora gostitelj nagradne igre vsakič postopati enako. Ne more enkrat ponuditi zamenjavo (npr. takrat, ko vidi, da nastopajoči kaže na vrata za katerimi se skriva avto), drugič pa ne (npr. takrat, ko nastopajoči kaže na vrata za katerimi je koza).

Vprašanja se je prijelo ime "*problem Montyja Halla*", po imenu voditelja popularne ameriške televizijske oddaje Pogodimo se (Let's Make a Deal),

v kateri je voditelj Monty Hall goste izzival, da so sprejemali ali zavračali najrazličnejše ponudbe, ki jim jih je zastavljal.

Marilyn je bralcu v svoji rubriki odgovorila, da se nam vrata vsekakor splača zamenjati, saj se tako verjetnost, da bomo zadeli avto, poveča za dvakrat. Tole je njen odgovor:

Seveda se splača zamenjati vrata.
Prva vrata imajo le 1/3 verjetnosti za zmago,
medtem ko imajo druga verjetnost 2/3.

Namig

Najlažje si vse skupaj predstavljate takole.

Predpostavimo, da je na voljo milijon vrat in vi izberete prva.

Nato voditelj, ki ve, kaj se nahaja za posameznimi vrati,
odpre vsa vrata razen vrat številka 777777.

V tem primeru bi zelo hitro zamenjali svoj izbor, kajne?

Se najinteligentnejša ženska na planetu moti?

Sledila je ploha kritik (več kot 10.000 pisem jeznih bralcev, med katerimi je bilo ogromno učiteljev matematike).

Skoraj 1000 pisem je bilo podpisanih z imeni (dr. nazivi, napisana na papirju z glavo katere od ameriških univerz - www.marilynvossavant.com).

Marylin bralce zavaja, saj se verjetnost za zadetek nikakor ne more spremeniti, če vmes zamenjamo izbor vrat.

Neki profesor matematike je bil zelo neposreden:
“Udarili ste mimo! ... Kot profesionalni matematik sem zelo zaskrbljen nad pomanjkanjem matematičnih veščin v širši javnosti. Prosim, da se opravičite in ste v prihodnosti bolj pazljivi.”

Drugi je Marylin celo obtožil, da je ona sama koza.

Polemika je pristala celo na naslovnici New York Timesa, v razpravo so se vključila tudi nekatera znana imena iz sveta matematike.

O odgovoru vos Savantove, da naj tekmovalec zamenja vrata, so razpravljali tako na hodnikih Cie kot v oporiščih vojaških pilotov ob Perzijskem zalivu. Analizirali so ga matematiki z MIT in računalniški programerji laboratorijev Los Alamos v Novi Mehiki.

Poleg žaljivih pisem, ki so njen odgovor kritizirala, je Marilyn vseeno prejela tudi nekaj pohval. Profesor s prestižnega MIT:

“Seveda imate prav. S kolegi v službi smo se poigrali s problemom in moram priznati, da je bila večina, med njimi sem bil tudi sam, sprva prepričana, da se motite!”

Eksperimentalna ugotovitev

Marilyn se kritik ni ustrašila - navsezadnje je objektivno izmerljivo po inteligenčnem količniku pametnejša od vseh svojih kritikov,

zato je v eni od svojih naslednjih kolumn vsem učiteljem v državi zadala nalogo, da to preprosto igrico igrajo s svojimi učenci v razredu (seveda ne s pravimi kozami in avtomobilom) in ji pošljejo svoje rezultate.

Te je nato tudi objavila in seveda so se povsem skladali z njenim nasvetom, da se v tem konkretnem primeru bistveno bolj splača spremeniti izbiro vrat.

Kdo ima prav?

Razprava o problemu Montyja Halla spada na področje, ki mu matematiki pravijo **pogojna verjetnost**.

Najbolj preprosto rečeno je to veda, ki se ukvarja s tem, kako prilagoditi verjetnost za posamezne dogodke, ko se pojavijo novi podatki.

Bistvo zapleta, ki je izzval tako obsežno in čustveno nabito reakcijo bralcev, je v tem, da so bralci večinoma spregledali ključni podatek.

Zelo pomembno je namreč dejstvo, da voditelj igre vnaprej ve, za katerimi vrati je avtomobil.

Ko v drugem delu odpre vrata, za katerimi se pokaže koza, vnaprej ve, da za temi vrati ni avtomobila.

Če voditelj te informacije ne bi imel in bi vrata odpiral povsem naključno tako kot igralec, se verjetnost za zadetek ob spremembi vrat res ne bi povečala.

Potem bi držale ugotovitve več 1000 bralcev, ki so poslali jezna pisma na uredništvo revije, da Marilyn ne pozna osnov matematike.

Matematična intuicija nam namreč pravi, da je verjetnost, da bo avto za enimi ali za drugimi vrati, ko so dvoja še zaprta, enaka.

To je seveda res, če zraven ne bi bilo še voditelja, ki ve več kot mi.

Najlažje nejasnost pojasnimo, če analiziramo dogajanje izza kulis, od koder ves čas vidimo, za katerimi vrati je avto in kje sta kozi.

Če tekmovalec že v prvo izbere vrata, za katerimi je avto, bo voditelj odprl katera koli od preostalih dveh vrat in zamenjava bo tekmovalcu v tem primeru le škodila.

Ampak to velja le za primer, če v prvo izbere vrata, za katerimi je avto, verjetnost za to pa je 1/3.

Če pa v prvo tekmovalec izbere vrata, za katerimi je koza, bo voditelj moral odpreti edina preostala vrata, za katerimi se nahaja koza.

V tem primeru se bo tekmovalcu zamenjava vrat v vsakem primeru obrestovala in bo tako z gotovostjo zadel avto.

Če v prvo tekmovalec izbere kozo, se mu vedno splača zamenjati, če pa v prvo izbere avto, se mu zamenjava ne izplača.

Verjetnost, da v prvo izbere kozo, je 2/3, medtem ko je verjetnost, da izbere avto, le 1/3.

Če se tekmovalec odloči za strategijo zamenjave, je zato verjetnost, da zadane avtomobil, 2/3, če zamenjavo zavrne, pa je verjetnost pol manjša, tj. 1/3.

Če se torej drži strategije zamenjave vrat, ko mu jo voditelj ponudi, bo tako vedno, ko v prvo izbere kozo, ob zamenjavi vrat dobil avto, kar ga do dobitka pripelje v $2\times$ večjem številu primerov, kot sicer. Verjetnost za zadetek se mu tako s 33% poveča na 66%.

Če vam ni takoj jasno, se ne sekirajte preveč. Tudi mnogi matematiki so potrebovali kar nekaj časa, da so si razjasnili ta problem.

Definicija pogojne verjetnosti

Opazujemo dogodek A ob poskusu X, ki je realizacija kompleksa pogojev K. Verjetnost dogodka A je tedaj $P(A)$.

Kompleksu pogojev K pridružimo mogoč dogodek B, tj. $P(B) > 0$. Realizacija tega kompleksa pogojev $K' = K \cap B$ je poskus X' in verjetnost dogodka A v tem poskusu je $P_B(A)$, ki se z verjetnostjo $P(A)$ ujema ali pa ne.

Pravimo, da je poskus X' poskus X s pogojem B in verjetnost $P_B(A)$ **pogojna verjetnost** dogodka A glede na dogodek B, kar zapišemo takole:

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

Pogojna verjetnost $P(A/B)$ v poskusu X' je verjetnost dogodka A v poskusu X s pogojem B.

Pogosto pogojno verjetnost pišejo tudi $P(A/B)$.

...Pogojna verjetnost

Denimo, da smo n–krat ponovili poskus X in da se je ob tem k_B–krat zgodil dogodek B. To pomeni, da smo v n ponovitvah poskusa X napravili k_B–krat poskus X'. Dogodek A se je zgodil ob poskusu X' le, če se je zgodil tudi B, t.j. $A \cap B$. Denimo, da se je dogodek $A \cap B$ zgodil ob ponovitvi poskusa $k_{A\cap B}$–krat. Potem je relativna frekvenca dogodka A v opravljenih ponovitvah poskusa X':

$$f_B(A) = f(A/B) = \frac{k_{A\cap B}}{k_B} = \frac{k_{A\cap B}/n}{k_B/n} = \frac{f(A\cap B)}{f(B)}$$

oziroma

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

Pogojna verjetnost P_B ima prav take lastnosti kot brezpogojna. Trojica $(B, \mathcal{D}_B, P_B)$, $\mathcal{D}_B = \{A \cap B \mid A \in \mathcal{D}\}$ je zopet verjetnostni prostor.

...Pogojna verjetnost

Primer: Denimo, da je v nekem naselju 900 polnoletnih prebivalcev. Zanima nas struktura prebivalcev po spolu (M – moški, Ž – ženski spol) in po zaposlenosti (Z – zaposlen(a), N – nezaposlen(a)). Podatke po obeh spremenljivkah uredimo v dvorazsežno frekvenčno porazdelitev, ki jo imenujemo tudi **kontingenčna tabela**:

$spol \setminus zap.$	Z	N	
M	460	40	500
Ž	240	160	400
	700	200	900

...Pogojna verjetnost

Poglejmo, kolikšna je verjetnost, da bo slučajno izbrana oseba moški pri pogoju, da je zaposlena.

$$P(Z) = \frac{700}{900} \quad , \quad P(M\cap Z) = \frac{460}{900}$$

$$P(M/Z) = \frac{P(M\cap Z)}{P(Z)} = \frac{460\cdot 900}{900\cdot 700} = \frac{460}{700}$$

ali neposredno iz kontingenčne tabele

$$P(M/Z) = \frac{460}{700}.$$

...Pogojna verjetnost

Iz formule za pogojno verjetnost sledi:

$$P(A\cap B) = P(B)\, P(A/B),$$

$$P(A\cap B) = P(A)\, P(B/A).$$

Torej velja:

$$P(A)\, P(B/A) = P(B)\, P(A/B).$$

Dogodka A in B sta **neodvisna**, če velja

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

Zato za neodvisna dogodka A in B velja $P(A\cap B) = P(A)\cdot P(B)$.

Za nezdružljiva dogodka A in B velja $P(A/B) = 0$.

. . . Pogojna verjetnost

Primer: Iz posode, v kateri imamo 8 belih in 2 rdeči krogli, $2\times$ na slepo izberemo po eno kroglo. Kolikšna je verjetnost dogodka, da je prva krogla bela (B_1) in druga rdeča (R_2).

1. Če po prvem izbiranju izvlečeno kroglo ne vrnemo v posodo (odvisnost), je:

$$P(B_1 \cap R_2) = P(B_1) \cdot P(R_2/B_1) =$$
$$= \frac{8}{10} \cdot \frac{2}{9} = 0,18.$$

2. Če po prvem izbiranju izvlečeno kroglo vrnemo v posodo (neodvisnost), je:

$$P(B_1 \cap R_2) = P(B_1) \cdot P(R_2/B_1) =$$
$$= P(B_1) \cdot P(R_2) = \frac{8}{10} \cdot \frac{2}{10} = 0,16.$$

. . . Pogojna verjetnost

Dogodka A in B sta neodvisna, če je $P(A/B) = P(A/\overline{B})$.

Nadalje velja

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/(A \cap B)).$$

(Tudi slednje pravilo lahko posplošimo naprej.)

Dogodki A_i, $i \in I$ so **neodvisni**, če je $P(A_j) = P(A_j / \bigcap_{i=1}^{j-1} A_i)$, $j \in I$.

Za neodvisne dogodke A_i, $i \in I$ velja

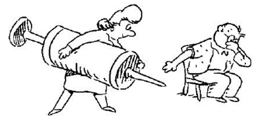
$$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i).$$

Naloga iz pogojne verjetnosti

Redko nalezljivo bolezen dobi ena oseba na 1000.

Imamo dober, a ne popoln test za to bolezen:

> *če ima neka oseba to bolezen, potem test to pokaže v 99% primerih, vendar pa test napačno označi tudi 2% zdravih pacientov za bolane.*



V Tvojem primeru je bil test pravkar **pozitiven**.

Kakšna je verjetnost, da si zares dobili nalezljivo bolezen?

. . . Naloga iz pogojne verjetnosti

Delamo z naslednjimi dogodki:

$\boldsymbol{A}$: pacient je dobil nalezljivo bolezen,

$\boldsymbol{B}$: pacientov test je bil pozitiven.

Izrazimo informacijo o učinkovitosti testov:

$\boldsymbol{P(A) = 0,001}$ (en pacient na 1000 se naleze),

$\boldsymbol{P(B/A) = 0,99}$ (test pravilno označi okuženega),

$\boldsymbol{P(B/\overline{A}) = 0,02}$ (test napačno označi zdravega).

Zanima nas $\boldsymbol{P(A/B)}$ (verjetnost, da smo se nalezli, če je test pozitiven).

Obrazec za razbitje in večstopenjski poskusi

Naj bo H_i, $i \in I$ **razbitje** gotovega dogodka: $\bigcup_{i \in I} H_i = G$, hkrati pa naj bodo dogodki paroma nezdružljivi: $H_i \cap H_j = N, i \neq j$.

Zanima nas verjetnost dogodka A, če poznamo verjetnost $P(H_i)$, in pogojno verjetnost $P(A/H_i)$ za $i \in I$:

$$A = A \cap (H_1 \cup H_2 \cdots H_n) = (A \cap H_1) \cup \cdots \cup (A \cap H_n).$$

Ker so tudi dogodki $A \cap H_i$ paroma nezdružljivi, velja:

$$P(A) = \sum_{i \in I} P(A \cap H_i) = \sum_{i \in I} P(H_i) P(A/H_i).$$

Na stvar lahko pogledamo tudi kot na večstopenjski poskus: v prvem koraku se zgodi natanko eden od dogodkov H_i, ki ga imenujemo hipoteza (hipoteze sestavljajo popoln sistem dogodkov).

Šele izidi na prejšnjih stopnjah določajo, kako bo potekal poskus na naslednji stopnji.

Omejimo se na poskus z dvema stopnjama.

Naj bo A eden izmed mogočih dogodkov na drugi stopnji. Včasih nas zanima po uspešnem izhodu tudi druge stopnje, verjetnost tega, da se je na prvi stopnji zgodil dogodek H_i.

Odgovor dobimo iz zgornjega obrazca in mu pravimo **Bayesov obrazec**:

$$P(H_k/A) = \frac{P(H_k)\cdot P(A/H_k)}{\sum_{i\in I} P(H_i)\cdot P(A/H_i)}.$$

Rev. T. Bayes

Leta 2001 je bila na vrsti že 300 letnica rojstva angleškega matematika Bayesa.

Zgled. Trije lovci so hkrati ustrelili na divjega prašiča in ga ubili. Ko so prišli do njega, so našli v njem eno samo kroglo. Kolikšne so verjetnosti, da je vepra ubil

(a) prvi,

(b) drugi,

(b) tretji

lovec, če poznamo njihove verjetnosti, da zadanejo: $0,2$; $0,4$ in $0,6$? Na ta način jim namreč lahko pomagamo pri pošteni delitvi plena (kajti ne smemo pozabiti, da imajo vsi v rokah nevarno orožje).

Sestavimo popoln sistem dogodkov in uporabimo dejstvo, da so lovci med seboj neodvisni, torej

$$P(A*B*C) = P(A)*P(B)*P(C).$$

To nam zna pomagati pri računanju verjetnosti hipotez.

	.2	.4	.6					
	prvi	drugi	tretji		P(H_i)	st.kr.	P(E/H_i)	P(E*H_i)
H1	1	1	1	,2*,4*,6	=0,048	3	0	0
H2	0	1	1	,8*,4*,6	=0,192	2	0	0
H3	1	0	1	,2*,6*,6	=0,072	2	0	0
H4	1	1	0	,2*,4*,4	=0,032	2	0	0
H5	1	0	0	,2*,6*,4	=0,048	1	1	0,048
H6	0	1	0	,8*,4*,4	=0,128	1	1	0,128
H7	0	0	1	,8*,6*,6	=0,288	1	1	0,288
H8	0	0	0	,8*,6*,4	=0,192	0	0	0
vsota					=1,000			0,464

$$P(\text{ena krogla je zadela}) = 0,048 + 0,128 + 0,288 = 0,464 = P(E).$$

Ostale verjetnosti računamo za preiskus:

$$\begin{aligned} P(\text{nobena krogla ni zadela}) &= 0,192 = P(N'),\\ P(\text{dve krogli sta zadeli}) = 0,192+0,072+0,032 &= 0,296 = P(D),\\ P(\text{tri krogle so zadele}) &= 0,048 = P(T).\end{aligned}$$

Vsota teh verjetnosti je seveda enaka 1.

Končno uporabimo Bayesov obrazec:

$$\begin{aligned} P(H_5/E) &= \frac{P(H_5*E)}{P(E)} = \frac{0,048}{0,464} = 0,103 &= P(\text{prvi je zadel}),\\ P(H_6/E) &= \frac{P(H_6*E)}{P(E)} = \frac{0,128}{0,464} = 0,276 &= P(\text{drugi je zadel}),\\ P(H_7/E) &= \frac{P(H_7*E)}{P(E)} = \frac{0,288}{0,464} = 0,621 &= P(\text{tretji je zadel}).\end{aligned}$$

Tudi vsota teh verjetnosti pa je enaka 1.
Delitev plena se opravi v razmerju 10,3 : 27,6 : 62,1 = 3 : 8 : 18 (in ne 2 : 4 : 6 oziroma 16,6 : 33,3 : 50, kot bi kdo utegnil na hitro pomisliti).

Bonus vprašanje:
Kako bi si razdelili plen, če bi v divjim prašiču našli dve krogli?

I.4. Bernoullijevo zaporedje neodvisnih poskusov

O zaporedju neodvisnih poskusov

$$X_1, X_2, \ldots, X_n, \ldots$$

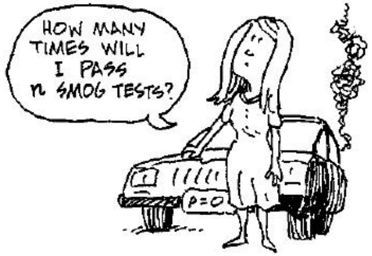
govorimo tedaj, ko so verjetnosti izidov v enem poskusu neodvisne od tega, kaj se zgodi v drugih poskusih.

Jakob Bernoulli um 1687

Zaporedje neodvisnih poskusov se imenuje **Bernoullijevo zaporedje**, *če se more zgoditi v vsakem poskusu iz zaporedja neodvisnih poskusov le dogodek A z verjetnostjo $P(A) = p$ ali dogodek $\overline{A}$ z verjetnostjo $P(\overline{A}) = 1 - P(A) = 1 - p = q$.*

Primer:

Primer Bernoullijevega zaporedja poskusov je met kocke, kjer ob vsaki ponovitvi poskusa pade šestica (dogodek A) z verjetnostjo $P(A) = p = 1/6$ *ali* ne pade šestica (dogodek $\overline{A}$) z verjetnostjo $P(\overline{A}) = 1 - p = q = 5/6$.



...Bernoullijevo zaporedje neodvisnih poskusov

V Bernoullijevem zaporedju neodvisnih poskusov nas zanima, kolikšna je verjetnost, da se v n zaporednih poskusih zgodi dogodek A natanko k–krat. To se lahko zgodi na primer tako, da se najprej zgodi k–krat dogodek A in nato v preostalih $(n-k)$ poskusih zgodi nasprotni dogodek $\overline{A}$:

$$P\Big(\bigcap_{i=1}^{k}(X_i = A) \cap \bigcap_{i=k+1}^{n}(X_i = \overline{A})\Big) = \prod_{i=1}^{k} P(A) \cdot \prod_{i=k+1}^{n} P(\overline{A}) = p^k \cdot q^{n-k}.$$

Dogodek $P_n(k)$, da se dogodek A v n zaporednih poskusih zgodi natanko k–krat, se lahko zgodi tudi na druge načine in sicer je teh toliko, na kolikor načinov lahko izberemo k poskusov iz n poskusov. Teh je $\binom{n}{k}$. Ker so ti načini nezdružljivi med seboj, je verjetnost dogodka $P_n(k)$ enaka

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Tej zvezi pravimo **Bernoullijev obrazec**.

Primer: Iz posode, v kateri imamo 8 belih in 2 rdeči krogli, na slepo izberemo po eno kroglo in po izbiranju izvlečeno kroglo vrnemo v posodo. Kolikšna je verjetnost, da v petih poskusih izberemo 3–krat belo kroglo?

Dogodek A je, da izvlečem belo kroglo. Potem je

$$p = P(A) = \frac{8}{10} = 0,8 \quad \text{in} \quad q = 1 - p = 1 - 0,8 = 0,2$$

Verjetnost, da v petih poskusih izberemo 3–krat belo kroglo, je:

$$P_5(3) = \binom{5}{3} 0,8^3 (1-0,8)^{5-3} = 0,205.$$

Računanje $P_n(k)$

Uporaba rekurzije: $P_n(0) = q^n$

$$P_n(k) = \frac{(n-k+1)p}{kq} P_n(k-1), \qquad \text{za } k = 1, \ldots$$

Stirlingov obrazec:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Poissonov obrazec: za majhne verjetnosti, tj. p blizu 0:

$$P_n(k) \approx \frac{(np)^k e^{-np}}{k!}.$$

Laplaceov točkovni obrazec:

$$P_n(k) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}.$$

Računanje $P_n(k)$

Program R: Vrednost $P_n(k)$ dobimo z ukazom

`dbinom(k,size=n,prob=p)`

```
> dbinom(50,size=1000,prob=0.05)
[1] 0.05778798
```

Izpeljava rekurzivne zveze

$$\frac{P_n(k)}{P_n(k-1)} = \frac{\binom{n}{k}p^k q^{n-k}}{\binom{n}{k-1}p^{k-1}q^{n-k+1}} =$$

$$= \frac{n!\,(k-1)!(n-k+1)!\,p}{k!(n-k)!\,n!\,q} = \frac{(n-k+1)p}{kq}$$

Torej je res

$$P_n(k) = \frac{(n-k+1)p}{kq} P_n(k-1), \qquad \text{za } k = 1, \ldots$$

Bernoullijev zakon velikih števil

IZREK 1 (J. Bernoulli, 1713) *Naj bo k frekvenca dogodka A v n neodvisnih ponovitvah danega poskusa, v katerem ima dogodek A verjetnost p. Tedaj za vsak $\varepsilon > 0$ velja*

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 1.$$

Ta izrek opravičuje statistično definicijo verjetnosti.