U dokazu teoreme jedinstvenosti za karakteristične funkcije koristimo Stonovu teoremu.

Teorema 1 (Stone) *Neka je $f(x)$ neprekidna funkcija na $[-l,l]$ i $f(-l)=f(l)$. Za $\forall\varepsilon>0$ postoji trigonometrijski polinom*

$$T_n(x)=\sum_{v=-n}^{n}a_v e^{i\pi v\frac{x}{l}}$$

takav da je $\sup_{[-l,l]}|T_n(x)-f(x)|<\varepsilon$.

Teorema 2 (Teorema jedinstvenosti) *Neka su F i G funkcije raspodjele koje imaju istu karakterističnu funkciju na $\mathbb{R}$ tj. za svako $t\in\mathbb{R}$ važi*

$$\int_{-\infty}^{\infty}e^{itx}dF(x)=\int_{-\infty}^{\infty}e^{itx}dG(x). \tag{1}$$

Tada je $F=G$.

Fiksirajmo $a,b\in\mathbb{R}\,(a<b),\varepsilon>0$ i formirajmo funkciju $f^{(\varepsilon)}$ za koju važi:

$$f^{(\varepsilon)}(x)=\begin{cases}0, & x<a-\varepsilon,\, x\geqslant b,\\ \frac{x-a+\varepsilon}{\varepsilon}, & a-\varepsilon\leqslant x<a,\\ 1, & a\leqslant x<b-\varepsilon,\\ \frac{b-x}{\varepsilon}, & b-\varepsilon\leqslant x<b.\end{cases}$$

Dokažimo da je

$$\int_{-\infty}^{\infty}f^{(\varepsilon)}(x)dF(x)=\int_{-\infty}^{\infty}f^{(\varepsilon)}(x)dG(x). \tag{2}$$

Neka je $n\in\mathbb{N}$ takav da je $[a-\varepsilon,b]\subset[-n,n]$ i neka je $(\delta_n,\,n\in\mathbb{N})$ takav niz da je $\delta_n\geqslant 0$ i $\delta_n\downarrow 0,\,n\to\infty$. Funkcija $f^{(\varepsilon)}$ ispunjava uslove Stonove teoreme te postoji konačna suma

$$f_n^{(\varepsilon)}(x)=\sum_k a_k e^{\frac{i\pi xk}{n}}$$

tako da važi

$$\sup_{x\in\mathbb{R}}\mid f^{(\varepsilon)}(x)-f_n^{(\varepsilon)}(x)\mid\leqslant\delta_n.$$

Funkciju $f_n^{(\varepsilon)}$ periodično produžimo na $\mathbb{R}$. Primijetimo

$$\sup_{x\in\mathbb{R}}\mid f_n^{(\varepsilon)}(x)\mid\leqslant 2.$$

Iz (1) slijedi

$$\int_{-\infty}^{\infty}f_n^{(\varepsilon)}(x)dF(x)=\int_{-\infty}^{\infty}f_n^{(\varepsilon)}(x)dG(x).$$

Imamo

$$\left|\int\limits_{-\infty}^{\infty} f^{(\varepsilon)}(x)dF(x) - \int\limits_{-\infty}^{\infty} f^{(\varepsilon)}(x)dG(x)\right| = \left|\int\limits_{-n}^{n} f^{(\varepsilon)}(x)dF(x) - \int\limits_{-n}^{n} f^{(\varepsilon)}(x)dG(x)\right| \leqslant$$

$$\left|\int\limits_{-n}^{n} f_n^{(\varepsilon)}(x)dF(x) - \int\limits_{-n}^{n} f_n^{(\varepsilon)}(x)dG(x)\right| + 2\delta_n \leqslant$$

$$\left|\int\limits_{-\infty}^{\infty} f_n^{(\varepsilon)}(x)dF(x) - \int\limits_{-\infty}^{\infty} f_n^{(\varepsilon)}(x)dG(x)\right| + 2\delta_n + 2F\big([-n,n]^c\big) + 2G\big([-n,n]^c\big), \tag{3}$$

gdje je $F(A) = \int_A dF(x)$ i $G(A) = \int_A dG(x)$. Puštajući da $n \to \infty$ zaključujemo da (3) teži ka 0. Ovim je dokazana jednakost (2).

Primijetimo $f^{(\varepsilon)}(x) \to I_{[a,b)}(x)$ kad $\varepsilon \to 0$. Na osnovu teoreme o dominantnoj konvergenciji iz (2) slijedi

$$\int\limits_{-\infty}^{\infty} I_{[a,b)}(x)dF(x) = \int\limits_{-\infty}^{\infty} I_{[a,b)}(x)dG(x)$$

tj. $F(b) - F(a) = G(b) - G(a)$. Zbog proizvoljnosti a i b zaključujemo da je $F(x) = G(x)$ za svako $x \in \mathbb{R}$.♦

Komentar. Teoremom jedinstvenosti je pokazana injektivnost korespondencije između funkcija raspodjele i karakterističnih funkcija.

1. Slučajne promjenljive $X_j, j = 1, 2, ...$ su nezavisne i jednako raspodijeljene sa raspodjelom

$$X_j : \begin{array}{cccc} 0 & 1 & ... & 9 \\ \frac{1}{10} & \frac{1}{10} & & \frac{1}{10} \end{array}.$$

Neka je

$$Y_n = \sum_{j=1}^{n} \frac{1}{10^j} X_j,\ n = 1, 2, ...$$

Dokazati da niz slučajnih promjenljivih

$$Y_n \xrightarrow{d} W : \mathcal{U}(0,1),\ n \to \infty.$$

$$f_{X_j}(t) = E^{itX_j} = \frac{1}{10}\sum_{k=0}^{9} e^{itk} = \frac{1}{10}\frac{\sin 5t}{\sin \frac{t}{2}} e^{i\frac{t}{2}}.$$

$$f_{Y_n}(t) = \frac{1}{10^n}\prod_{j=1}^{n} \frac{\sin 5\frac{t}{10^j}}{\sin \frac{t}{2\cdot 10^j}} e^{\frac{9it}{2\cdot 10^j}} = \frac{1}{10^n}\frac{\sin \frac{t}{2}}{\sin \frac{t}{2\cdot 10^n}} e^{\frac{it}{2}(1-10^{-n})} \to 2\frac{\sin \frac{t}{2}}{t} e^{i\frac{t}{2}} = f_W(t).$$

Koristili $e^{i\varphi}-1=2i\sin\frac{\varphi}{2}e^{i\frac{\varphi}{2}}$; u proizvodu sinusa nakon što sinuse eksplicitno zapišemo dolazi do skraćivanja.

2. Neka je $X_n, n=1,2,...$ niz slučajnih promjenljivih sa $\mathcal{U}((-4-\frac{1}{n},-4+\frac{3}{n})\cup(0,\frac{2}{n}))$ raspodjelom. Metodom karakterističnih funkcija ispitati konvergenciju u raspodjeli datog niza.

$$f_{X_n}(t)=\int\limits_{-4-\frac{1}{n}}^{-4+\frac{3}{n}}\frac{n}{6}e^{itx}dx+\int\limits_{0}^{\frac{2}{n}}\frac{n}{6}e^{itx}dx\to\frac{2}{3}e^{-4it}+\frac{1}{3}=f_W(t),\ W:\ \begin{matrix}-4 & -0\\ \frac{2}{3} & \frac{1}{3}\end{matrix}.$$

Teorema 3 **Kolmogorovljeva nejednakost.** *Neka su $X_k, k=1,2,...,n$, nezavisne slučajne promjenljive, $EX_k=0, EX_k^2<\infty, S_k=X_1+...+X_k$. Tada za svako $\varepsilon>0$ važi*

$$P\{\max_{1\leqslant k\leqslant n}\mid S_k\mid\geqslant\varepsilon\}\leqslant\frac{DS_n}{\varepsilon^2}.$$

Označimo

$$A=\{\max_{1\leqslant k\leqslant n}\mid S_k\mid\geqslant\varepsilon\}; A_k=\{\mid S_i\mid<\varepsilon, i=1,...,k-1,\mid S_k\mid\geqslant\varepsilon\}, 1\leqslant k\leqslant n.$$

Konstatujmo,

$$A=\sum_{k=1}^{n}A_k; ES_n^2\geqslant ES_n^2I_A=\sum_{k=1}^{n}ES_n^2I_{A_k}.$$

$$ES_n^2I_{A_k}=E(S_k+(X_{k+1}+...+X_n))^2I_{A_k}=ES_k^2I_{A_k}+$$
$$2ES_k(X_{k+1}+...+X_n)I_{A_k}+E(X_{k+1}+...+X_n)^2I_{A_k}\geqslant ES_k^2I_{A_k}.$$

Naime $ES_kI_{A_k}(X_{k+1}+...+X_n)=ES_kI_{A_k}E(X_{k+1}+...+X_n)=0$, koristimo nezavisnost promjenljivih $S_kI_{A_k}$ i $X_{k+1}+...+X_n$. Na kraju,

$$ES_n^2\geqslant\sum_{k=1}^{n}ES_k^2I_{A_k}\geqslant\varepsilon^2\sum_{k=1}^{n}P(A_k)=\varepsilon^2P(A).$$

Posljedica 1 *Neka su $X_k, k=1,2,...,n$, nezavisne slučajne promjenljive sa konačnim disperzijama $DX_k, S_k=X_1+...+X_k$. Tada za svako $\varepsilon>0$ važi*

$$P\{\max_{1\leqslant k\leqslant n}\mid S_k-ES_k\mid\geqslant\varepsilon\}\leqslant\frac{DS_n}{\varepsilon^2}.$$

Ako su $w_k=X_k-EX_k$ i $S'_k=w_1+...+w_k$, tada je $Ew_k=0, DS'_k=DS_k$. Imamo na osnovu upravo dokazane teoreme

$$P\{\max_{1\leqslant k\leqslant n}\mid S_k-ES_k\mid\geqslant\varepsilon\}=P\{\max_{1\leqslant k\leqslant n}\mid S'_k\mid\geqslant\varepsilon\}\leqslant\frac{DS'_n}{\varepsilon^2}=\frac{DS_n}{\varepsilon^2}.$$

Teorema 4 **Kolmogorovljev zakon velikh brojeva.** *Neka je* $X_n, n=1,2,...,$ *niz nezavisnih slučajnih promjenljivih,* $EX_n = 0$ *i red* $\sum\limits_{n=1}^{\infty}\frac{DX_n}{n^2}$ *konvergira. Tada* $\frac{1}{n}\sum\limits_{i=1}^{n}X_i \xrightarrow{si} 0, n\to\infty$*, tj. za niz* X_n *važi strogi zakon velikih brojeva.*

Neka je $Y_n = \max\limits_{1\leqslant l\leqslant 2^n} \mid \sum\limits_{j=1}^{l} X_j \mid$. Za $k, 2^{n-1} < k \leqslant 2^n$ važi

$$\mid \frac{1}{k}\sum_{j=1}^{k}X_j \mid \leqslant \frac{1}{2^{n-1}}\max_{1\leqslant l\leqslant 2^n}\mid \sum_{j=1}^{l}X_j \mid = 2\frac{Y_n}{2^n}.$$

Za dokaz tvrđenja iz teoreme dovoljno je pokazati da $\frac{Y_n}{2^n}\xrightarrow{si} 0$, teorema o policajcima. Neka je $\varepsilon > 0$. Na osnovu Kolmogorovljeve nejednakosti dobijamo

$$P\Big\{\frac{Y_n}{2^n}\geqslant\varepsilon\Big\}\leqslant\frac{1}{4^n\varepsilon^2}\sum_{l=1}^{2^n}DX_l$$

te je

$$\sum_{n=1}^{\infty}P\Big\{\frac{Y_n}{2^n}\geqslant\varepsilon\Big\}\leqslant\frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{4^n}\sum_{l=1}^{2^n}DX_l\leqslant\frac{1}{\varepsilon^2}\sum_{l=1}^{\infty}DX_l\sum_{l\leqslant 2^n}\frac{1}{4^n}\leqslant\frac{1}{\varepsilon^2}\sum_{l=1}^{\infty}\frac{1}{1-\frac{1}{4}}\frac{DX_l}{l^2}<\infty$$

Dakle, $\frac{Y_n}{2^n}\xrightarrow{si} 0$.

Posljedica 2 *Neka je* $X_n, n=1,2,...,$ *niz nezavisnih slučajnih promjenljivih i neka red* $\sum\limits_{n=1}^{\infty}\frac{DX_n}{n^2}$ *konvergira. Tada* $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - EX_i)\xrightarrow{si} 0$*, tj. za niz* X_n *važi strogi zakon velikih brojeva.*

Formirajmo niz $w_n = X_n - EX_n$. Očigledno, $Ew_n = 0$, $Dw_n = DX_n$ i red $\sum\limits_{n=1}^{\infty}\frac{Dw_n}{n^2} = \sum\limits_{n=1}^{\infty}\frac{DX_n}{n^2}$ na osnovu pretpostavke konvergira. Na osnovu upravo dokazane teoreme slijedi $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - EX_i) = \frac{1}{n}\sum\limits_{i=1}^{n}w_i \xrightarrow{si} 0$.

Posljedica 3 **Borelov zakon velikh brojeva.** *U Bernulijevoj shemi* $\frac{S_n}{n}\xrightarrow{si} p$.

Znamo da je $S_n = \sum\limits_{k=1}^{n} I_{A_k}$, gdje je A_k događaj da se u k tom opitu ostvari događaj A tj. uspjeh. Promjenljive iz niza I_{A_k} su nezavisne, $EI_{A_k} = p$, $DI_{A_k} = p(1-p)$ te iz Kolmogorovljevog zakona velikih brojeva slijedi tvrđenje.

Posljedica 4 *Neka je X_n niz nezavisni jednako raspodijeljenih slučajnih promjenljivih sa konačnom disperzijom i neka je $EX_n = a, S_n = X_1 + ... + X_n$. Tada za niz X_n važi strogi zakon velikih brojeva tj. $\frac{S_n}{n} \stackrel{si}{\to} a$.*

Teorema 5 **Kolmogorovljev zakon velikih brojeva.** *Neka je X_n niz nezavisnih jednako raspodijeljenih slučajnih promjenljivih, $S_n = X_1 + ... + X_n$. $\frac{S_n}{n} \stackrel{si}{\to} a$ ako i samo ako je $EX_n = a$.*

Statistički eksperiment podrazumijeva mjerenja obilježja. Mi ćemo se uglvnom baviti nezavisnim mjerenjima. Rezultati mjerenja se tretiraju kao realizacije nezavisnih slučajnih promjenljivih $X_1, X_2, ..., X_n$ i svaka slučajna promjenljiva je raspodijeljena kao obilježje X.

Definicija 1 *Neka obilježje X ima funkciju raspodjele F. Prost slučajni uzorak obima n iz raspodjele F je slučajni vektor $(X_1, X_2, ..., X_n)$ gdje su $X_1, X_2, ..., X_n$ nezavisne slučajne promjenljive sa istom funkcijom raspodjele F.*

Prost slučajni uzorak je matematički model n nezavisnih mjerenja obilježja. Često se na osnovu predznanja o izučavanom fenomenu može pretpostaviti da raspodjela obilježja pripada nekoj familiji $\mathfrak{R}$ koju nazivamo familijom dopustivih raspodjela. Zadatak matematičke statistike je da se na osnovu registrovanih podataka zaključi što je moguće više o nepoznatoj funkciji raspodjele obilježja, a što je ekvivalentno sa tim da se sazna što je moguće više o nepoznatoj raspodjeli obilježja. Ubuduće ćemo umjesto prost slučajni zorak govoriti uzorak.

Primjer 1 Registrujemo α čestice koje emituje grumen radioaktivne materije, intenzitet potoka čestica je λ, λ je nepoznato. Slučajna promjenljiva X_1 predstavlja broj čestica koje se emituju u toku prvog minuta, X_2 predstavlja broj čestica koje se emituju u toku drugog minuta,...,X_n predstavlja broj čestica koje se emituju u toku n-tog minuta. Obilježje X je broj čestica koje se emituju u toku jednog minuta i X ima $\mathcal{P}(\lambda)$ raspodjelu. Slučajne promjenljive $X_1, ..., X_n$ su nezavisne i svaka ima $\mathcal{P}(\lambda)$ raspodjelu. $(X_1, ..., X_n)$ je prost slučajan uzorak. Koristili smo činjenice 1. slučajne promjenljive koje prebrojavaju čestice na disjunktnim vremenskim intervalima su nezavisne 2. raspodjela slučajne promjenljive koja prebrojava čestice na vremenskom intervalu dužine t je $\mathcal{P}(\lambda t)$. U ovom primjeru familiju dopustivih raspodjela čine Puasonove raspodjele sa parametrom $\lambda, \lambda > 0$. Skup $\Lambda = (0, \infty)$ mogućih vrijednosti za λ nazivamo parametarskim skupom. Ako imamo neku dodatnu informaciju o λ, recimo da je $1 < \lambda < 2$ tada je $\Lambda = (1, 2)$. U ovom modelu označimo sa W_1 vrijeme do emitovanja prve čestice, sa W_2 vrijeme od emitovanja prve do emitovanja druge, sa W_3 vrijeme od emitovanja deruge do emitovanja treće, sa W_n vrijeme od emitovanja $n-1$-e do emitovanja n-te. Pomenuta vremena su slučajna te su $W_1, ..., W_n$ slučajne promjenljive i one su nezavisne, $(W_1, ..., W_n)$ je prost slučajni uzorak, a odgovarajuća familija raspodjela je $\mathcal{E}(\lambda), \lambda > 0$. Teorijsko objašnjenje zbog čega se u modelu pojavljuju Puasonova i eksponencijalna raspodjela te zbog čega su promjenljive $X_1, ..., X_n$ odnosno $W_1, ..., W_n$ nezavisne biće dato u okviru predmeta Slučajni procesi.

Primjer 2. Obilježje X je rezultat mjerenja fizičke veličine, dok rezultati n mjerenja predstavljaju prost slučajni uzorak (uzorak). Gaus je tvorac teorije grešaka kod mjerenja fizičkih veličina. Primijetio ja da rezultati mjerenja udaljenosti konkretnog nebeskog tijela od Zemlje variraju kao posljedica postojanja greške mjerenja. Greška mjerenja ε se dobija sumiranjem malih grešaka uzrokovanim faktorima: temperaturna kolebanja koja utiču na mjerni instrument, intenzitet sunčevih zraka koji može dovesti do savijanja instrumenta, vazdušna strujanja, vibracije i mnogim drugim. Sumarna greška ε na osnovu CGT (ako ne postoji sistematska greška) ima $\mathcal{N}(0, \sigma^2), m > 0$ raspodjelu. Ako je stvarna udaljenost d tada obilježje X koje predstavlja rezultat mjerenja ima $\mathcal{N}(m, \sigma^2)$ raspodjelu.

Primjer 3. Obilježje X je rezultat mjerenja koeficijenta inteligencije osobe, dok rezultati mjerenja koeficijenta inteligencije n osoba predstavljaju prost slučajni uzorak (uzorak). Familija dopustivih raspodjela je normalna.

Primjer 4. Obilježje je kilometraža koju obezbjeđuje jedna baterija automobilu na električni pogon. Familija dopustivih raspodjela je eksponencijalna.

Primjer 5. Obilježje X je 1 ako osoba koja ima pravo glasa, glasa za kandidata A, a 0 ako ne glasa za kandidata A. Raspodjela obilježja X je

$$X : \begin{array}{cc} 0 & 1 \\ 1-p=q & p \end{array}$$

i p je relativna učestalost onih koji glasaju za kandidata A. Drugim riječima p je vjerovatnoća da slučajno izbrana osoba sa pravom glasa, glasa za A. Ako osobe sa pravom glasa u uzorak biramo metodom bez vraćanjanja, tada su promjenljive $X_1, ..., X_n$ zavisne i raspodijeljene kao i obilježje.

Definicija 2 *Neka je* $f : \mathrm{R}^n \to \mathrm{R}$ *Borelova funkcija i* $(X_1, ..., X_n)$ *uzorak. Slučajna promjenljiva* $V = f(X_1, ..., X_n)$ *se naziva statistika.*

Neki autori statistiku izjednačavaju sa funkcijom f. Realizaciju uzorka $(X_1, ..., X_n)$ označavamo sa $(x_1, ..., x_n)$ i nazivamo realizovanim uzorkom. Realizovana statistika je $v = f(x_1, ..., x_n)$.

Primjeri statistika.

Uzoračka sredina. $f(x_1, ..., x_n) = \frac{X_1+...+X_n}{n}, \overline{X}_n = \frac{X_1+...+X_n}{n}$. Neka je $EX = a, DX = \sigma^2$. Tada je $E\overline{X}_n = a, D\overline{X}_n = \frac{\sigma^2}{n}$.

Uzoračka disperzija. $\overline{S}_n^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - \overline{X}_n^2; E\overline{S}_n^2 = EX^2 - E\overline{X}_n^2 = EX^2 - \frac{1}{n}EX^2 - \frac{n^2-n}{n^2}(EX)^2 = \frac{n-1}{n}\sigma^2$.

Popravljena uzoračka disperzija. $\hat{S}_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}_n^2; E\hat{S}_n^2 = \sigma^2$.

Uzorački minimum. $Y_1 = \min\limits_{1\leqslant i\leqslant n} X_i$. **Uzorački maksimum.** $Y_n = \max\limits_{1\leqslant i\leqslant n} X_i$.

Primjer. U kutiji se nalazi N listica i na njima su zapisani brojevi $a_1, a_2, ..., a_N$. Obilježje je broj na listici.

a) model sa vraćanjem, vadimo n listica.

$P\{X_i = a_j\} = \frac{1}{N}, i = 1, 2, ..., n; j = 1, 2, ..., N$, $\frac{a_1+...+a_N}{N} = a = EX_i$, $DX_i = \sigma^2 = \frac{a_1^2+...+a_N^2}{N} - a^2$, $E\overline{X}_n = a$, $D\overline{X}_n = \frac{\sigma^2}{n}$.

b) model bez vraćanja, $n \leqslant N$.

$P\{X_i^{'} = a_j\} = \frac{1}{N}, i = 1,2,...,n; j = 1,2,...,N$, $a = EX_i^{'}$, $DX_i^{'} = \sigma^2$, $\rho_{X_k^{'},X_l^{'}} = -\frac{1}{N-1}$, $E\overline{X}_n^{'} = a$, $D\overline{X}_n^{'} = \frac{\sigma^2}{n}\frac{N-n}{N-1}$, $D\overline{X}_n^{'} < D\overline{X}_n$ $n \geqslant 2$.

Primjer. U kutiji se nalazi M bijelih i $N-M$ crnih kuglica. Ocjenjujemo relativnu učestalost bijelih kuglica to jest $\frac{M}{N}$. Ono što ćemo uraditi istovremeno obuhvata slučajeve a) M nepoznato, N poznato i b) M nepoznato i N nepoznato.

1. Model sa vraćanjem, vadimo n kuglica.

$$X_i : \begin{matrix} 0 & 1 \\ 1-\frac{M}{N} & \frac{M}{N} \end{matrix}, i = 1,2,...,n, \sum_{i=1}^{n} X_i : \mathcal{B}(n, \frac{M}{N}),\ E\overline{X}_n = \frac{M}{N},\ D\overline{X}_n = \frac{M}{nN}(1-\frac{M}{N}).$$

U istu shemu se uklapa model u kome ocjenjujemo nepoznatu vjerovatnoću padanja pisma. Novčić bacamo n puta.

$$X_i : \begin{matrix} 0 & 1 \\ 1-p & p \end{matrix}, i = 1,2,...,n, \sum_{i=1}^{n} X_i : \mathcal{B}(n,p),\ E\overline{X}_n = p,\ D\overline{X}_n = \frac{p(1-p)}{n}.$$

2. Model bez vraćanja, vadimo n kuglica, $n \leqslant N$.

$$X_i^{'} : \begin{matrix} 0 & 1 \\ 1-\frac{M}{N} & \frac{M}{N} \end{matrix}, i = 1,2,...,n,\ \rho_{X_k^{'},X_l^{'}} = \frac{\frac{n-1}{N-1} - \frac{n}{N}}{1-\frac{n}{N}},\ E\overline{X}_n^{'} = \frac{M}{N},\ D\overline{X}_n^{'} = \frac{M}{nN}(1-\frac{M}{N})\frac{N-n}{N-1}.$$

Primjer Ocjenjujemo nepoznatu učestalost osoba čiji je koeficijent inteligencije između 105 i 115. $P\{105 < X < 115\} = p$, p je nepoznato. Zbog nezavisnosti i jednake raspodijeljenosti promjenljivih $X_1,..,X_n$ događaji $A_1 = \{105 < X_1 < 115\},...,A_n = \{105 < X_n < 115\}$ su nezavisni. Formiranjem slučajnih promjenljivih $X_1 = I_{A_1},...,X_n = I_{A_n}$ problem ocjene nepoznatog p (kao i formiranja odgovarajućeg intervala povjerenja) se rješava korišćenjem statistike $\overline{X}_n$.

Veza modela sa vraćanjem i bez vraćanja Statistika koja prebrojava bijele kuglice u uzorku obima n (model bez vraćanja) ima hipergeometrijsku raspodjelu. Označimo tu statistiku sa W. Razmotrimo granični slučaj u kome $\frac{M}{N} \to p$. Lako se provjerava

$$P\{W = k\} \to \binom{n}{k} p^k (1-p)^{n-k}, k = 0,1,...,n.$$

Dakle, u graničnom slučaju, statistika W ima $\mathcal{B}(n,p)$ raspodjelu. Budući da statistika $\sum_{i=1}^{n} X_i$ iz modela 1 ima $\mathcal{B}(n, \frac{M}{N})$ raspodjelu, dobijeni granični rezultat u praksi koristimo na sljedeći način: Ako je broj članova populacije N velik, uzoračke sredine tj. srednji broj izvučenih bijelih kuglica u oba modela imaju istu raspodjelu (preciznije približno istu, ali razlika je zanemarljiva). Ovo zapažanje je značajno kod formiranja intervala povjerenja nepoznate učestalosti. Naime, podatke biramo po modelu bez vraćanja. a kad formiramo interval povjerenja, koristimo rezultat dobijen za model sa vraćanjem. Na primjer, ocjenjujemo učestalost pušača.

χ^2 **raspodjela sa** n **stepeni slobode.** Neka su $X_1,...,X_n$ nezavisne slučajne promjenljive, svaka promjenljiva ima $\mathcal{N}(0,1)$ raspodjelu. Raspodjela slučajne promjenljive $\chi_n^2 = \sum\limits_{i=1}^{n} X_i^2$ se naziva χ^2 sa n stepeni slobode. Potražimo odgovarajuću gustinu raspodjele, označimo je sa g_n. Označimo sa $V_n(x)$ zapreminu lopte $x_1^2 + ... + x_n^2 < x, x > 0$. Iz analize znamo da je $V_n(x) = x^{\frac{n}{2}} V_n(1)$. Znamo da je $(t+h)^{\frac{n}{2}} - t^{\frac{n}{2}} = \frac{n}{2} t^{\frac{n}{2}-1} h + o(h)$.

$$P\{t \leqslant \chi_n^2 < t+h\} = \int_{t \leqslant \sum\limits_{i=1}^{n} < t+h} \cdots \int (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum\limits_{i=1}^{n} x_i^2} dx_1 ... dx_n = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(t+\theta h)} [V_n(t+h) - V_n(t)],$$

$$\Rightarrow g_n(t) = \frac{n}{2}(2\pi)^{-\frac{n}{2}} e^{-\frac{t}{2}} t^{\frac{n}{2}-1} V_n(1), \text{ te iz } \int_0^{\infty} g_n(t)dt = 1 \Rightarrow g_n(t) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}, t > 0.$$

Konstatujmo da je $E\chi_n^2 = n; D\chi_n^2 = 2n, f_n(t) = (1-2it)^{-\frac{n}{2}}$.

Studentova raspodjela sa n stepeni slobode. Neka su $X_0, X_1, X_2, ..., X_n$ nezavisne slučajne promjenljive, svaka ima $\mathcal{N}(0,1)$ raspodjelu. Za slučajnu promjenljivu $t_n = X_0 \left(\frac{\sum\limits_{i=1}^{n} X_i^2}{n} \right)^{-\frac{1}{2}}$ kažemo da ima Studentovu raspodjelu sa n stepeni slobode. Gustina $\psi_n(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, t \in \mathbb{R}$. Konstatujmo, $Et_n = 0, n > 1; Dt_n =: \begin{cases} \infty, & n \leqslant 2 \\ \frac{n}{n-2}, & n > 2. \end{cases}$

TEOREMA 6 **Fišerova teorema.** *Neka obilježje X ima $\mathcal{N}(m, \sigma^2)$ raspodjelu. Tada važi:*
a) $\frac{\overline{X}_n - m}{\sigma} \sqrt{n} : \mathcal{N}(0,1)$.
b) Statistike $\overline{X}_n$ i $\overline{S}_n^2$ su nezavisne.
c) $\frac{n\overline{S}_n^2}{\sigma^2} : \chi_{n-1}^2$ *i* $\frac{(n-1)\hat{S}_n^2}{\sigma^2} : \chi_{n-1}^2$
d) $\frac{\overline{X}_n - m}{\hat{S}_n} \sqrt{n} : t_{n-1}$.

Glivenko-Kantelijeva centralna teorema matematičke statistike.

DEFINICIJA 3 *Neka je $(X_1, ..., X_n)$ uzorak iz raspodjele F. Funkcija*

$$F_n(x, \omega) = F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I\{X_k(\omega) < x\},\ x \in \mathbb{R}$$

se naziva empirijska funkcija raspodjele.

Ako se fiksira ω dobija se funkcija po x i ona je jednaka relativnoj učestalosti izmjerenih podataka manjih od x. Pretpostavimo da su izmjereni podaci $x_1, x_2, ..., x_n$ (tj. realizovani uzorak je je $(x_1, ..., x_n)$) i da su različiti. Imamo

$$F_n(x) = \begin{cases} 0, & \text{ako je } x \leqslant x_{(1)}, \\ \frac{k}{n}, & \text{ako je } x_{(k)} < x \leqslant x_{(k+1)},\ k = 1, 2, ..., n-1, \\ 1, & \text{ako je } x > x_{(n)}. \end{cases}$$

Primijetimo da je funkcija $F_n(x)$ funkcija raspodjele slučajne promjenljive koja ima ravnomjernu diskretnu raspodjelu na skupu $\{x_1, x_2, ..., x_n\}$.

Ako fiksiramo x tada je $F_n(x,\omega)$ slučajna promjenljiva. Primijetimo, $\sum_{k=1}^{n} I\{X_k(\omega) < x\}$ je slučajna promjenljiva sa $\mathcal{B}(n, F(x))$ raspodjelom te je

$$P\Big\{F_n(x,\omega) = \frac{k}{n}\Big\} = \binom{n}{k} F^k(x)(1 - F(x))^{n-k},\ k = 0, 1, ..., n.$$

$F_n(x,\omega)$ je slučajni proces. Funkcija F se naziva teorijska funkcija raspodjele obilježja X.

Empirijska funkcija raspodjele u slučaju kada je obim uzorka veliki aproksimira teorijsku funkciju raspodjele. Preciznije, važe sljedeće dvije teoreme.

Teorema 7 *Za svaki realni broj x važi*

$$P\{\lim_{n\to\infty} F_n(x) = F(x)\} = 1.$$

Tvrđenje teoreme je neposredna posljedica jakog zakona velikih brojeva.

Teorema 8 **Glivenko-Kantelijeva centralna teorema matematičke statistike.** *Neka je F teorijska funkcija raspodjele obilježja X. Važi*

$$P\Big\{\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \mid F_n(x) - F(x) \mid = 0\Big\} = 1.$$

Drugim riječima, skoro izvjesno $F_n(x) \rightrightarrows F(x)$. Neka je $A = \{\omega : \sup_{x\in\mathbb{R}} \mid F_n(x,\omega) - F(x) \mid = 0\}$. Glivenko-Kantelijeva teorema tvrdi da je vjerovatnoća-mjera skupa A jednaka 1. Neka je $\omega_0 \in A$. Tada za $\forall\varepsilon > 0$, $\exists n_0 = n(\omega_0, \varepsilon)$ takav da za $\forall n > n_0$ i $\forall x \in \mathbb{R}$ važi $\mid F_n(x,\omega_0) - F(x) \mid < \varepsilon$.