PRELEGEREA VII
STATISTICĂ MATEMATICĂ

I. Alte repartiţii teoretice

7.1. Repartiţia hipergeometrică

În cazul legii hipergeometrice efectivul populaţiei (**N**) este foarte mare în raport cu numărul de obiecte extrase din urnă (**n**), iar extragere este una exhaustivă, adică obiectul extras din urnă nu se mai pune înapoi. La plecare cele două categorii de obiecte sînt în proporţie de **p** şi **q**, unde **p + q = 1**. Pe măsură ce se execută extragerile succesive, cele două proporţii se schimbă.

Cazuri posibile: - dimensiunea populaţiei este **N**;
- o extragere cuprinde **n** elemente.

Cazuri favorabile: - există N_p elemente în populaţie din categoria (**1**);
- există **k** elemente de tipul (**1**) într-o extragere;
- există $C_{N_p}^{k}$ moduri de extragere;
- există N_q elemente în populaţie din categoria (**2**);
- trebuie să fie **n - k** elemente de tipul (**2**) într-o extragere;
- există $C_{N_q}^{n-k}$ moduri de extragere.

Numărul total de cazuri favorabile este: $C_{N_p}^{k} \cdot C_{N_q}^{n-k}$.

Probabilitatea de a avea **k** obiecte extrase din prima categorie, care la plecare era în proporţie de **p**, este dată de relaţia:

P(k) = număr_cazuri_favorabile/ număr_cazuri_posibile = $C_{N_p}^{k} \cdot C_{N_q}^{n-k} / C_{N}^{n}$.

Proprietăţi.

1. $M(X) = \mu = np$;

2. $D^2(X) = \sigma^2 = \mu_2 = npq \dfrac{N-n}{N-1}$;

3. Mo este partea întreagă a expresiei: $(n+1)\dfrac{N.p+1}{N+2}$;

Observaţie. Dacă **n** -> ∞, legea hipergeometrică tinde către legea binomială. A se vedea observaţia **2** de la legea binomială.

Aplicaţie. O urnă conţine **4** bile negre şi **6** albe. Extrăgînd simultan **3** bile, care este probabilitatea de a avea o bilă neagră şi doua albe?

Se aplică legea hipergeometrică: $P(1) = \dfrac{C_4^1 . C_6^2}{C_{10}^3} = 0.50$.

7.2. Repartiţia Poisson

Repartiţia Poisson (**legea probabilităţilor mici** sau **legea evenimentelor rare**) este asemănătoare cu cea binomială, cu deosebirea că **n** -> ∞, **p** -> **0** şi produsul **np = a** este constant. Altfel spus, repartiţia Poisson este un caz limită al repartiţiei binomiale.

Înlocuind $\mathbf{p}=\frac{a}{n}$ în expresia legii binomiale, $\mathbf{P_n(k)}=C_n^k\,.\,\mathbf{p^k.q^{n-k}}$ rezultă: $\mathbf{P_n(k)}$ $=\frac{n(n-1)...(n-k+1)}{k!}\cdot\left[\frac{a}{n}\right]^k\cdot\left[1-\frac{a}{n}\right]^{n-k}$ sau, în mod echivalent, $\mathbf{P_n(k)}$ = $\frac{a^k}{k!}\cdot\left[\frac{n(n-1)...(n-k+1)}{n^k}\right]\cdot\frac{\left[1-\frac{a}{n}\right]^n}{\left[1-\frac{a}{n}\right]^k}$. Împărţinînd fiecare factor din prima paranteză dreaptă prin **n**, se poate scrie:

$$\mathbf{P_n(k)} = \frac{a^k}{k!}\cdot\left[1.(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{k-1}{n})\right]\cdot\frac{\left[1-\frac{a}{n}\right]^n}{\left[1-\frac{a}{n}\right]^k}.$$

Trecînd la limită şi ţinînd cont că limita expresiilor $\left[1-\frac{a}{n}\right]^n$ şi $\left[1-\frac{a}{n}\right]^k$ tinde la $\mathbf{e^{-a}}$(deoarece se poate scrie: $\left[1-\frac{a}{n}\right]^n=\left[(1-\frac{a}{n})^{-\frac{n}{a}}\right]^{-a}$) şi respectiv **1**, iar factorii din prima paranteză pătrată tind la **1**, pentru **n -> ∞**, se obţine:

$$\boldsymbol{\psi_n(k)} = \lim_{n->\infty}\mathbf{P_n(k)} = \frac{a^k e^{-a}}{k!}.$$

Se reaminteşte că **e = 2.7182818284…** (**constanta lui Euler**) şi **e = lim(1+$\frac{1}{n}$)$^{\mathbf{n}}$** pentru **n -> ∞**.

Se observă că, repartiţia Poisson este caracterizată numai de constanta **a** şi este generată de probabilităţile pentru diferite valori ale lui **k**, iar funcţia de repartiţie se calculează prin însumarea lor:

$$\mathbf{F_n(x)} = \sum_{k=0}^{x}\Psi_n(k).$$

Observaţie. Cu o eroare neglijabilă, pentru cazurile **n ≥ 50** şi **p ≤ 0.10**, se poate utiliza legea Poisson în locul legii binomiale, deoarece modalitatea de calcul este mult simplificată.

Proprietăţi.

1. M(X) = μ = a**;**

$$\mathbf{M(X)} = \sum_{k=0}^{\infty}k.\Psi_n(k) = \sum_{k=0}^{\infty}\mathbf{k.}\frac{a^k e^{-a}}{k!} = \sum_{k=1}^{\infty}\mathbf{.k.e^{-a}.}\frac{a^k}{k!} =$$

$\mathbf{a.}\,e^{-a}\sum_{k=1}^{\infty}\frac{a^{k-1}}{(k-1)!}$, deoarece primul termen al seriei este egal cu zero pentru **k = 0**, simplificînd prin **k** şi rearanjînd ordinea factorilor în termenul general al sumei.

Se poate demonstra că suma seriei $\sum_{k=1}^{\infty} \frac{a^{k-1}}{(k-1)!}$ = $\mathbf{e^a}$.

Deci, rezultă că $\mathbf{M(X) = a \cdot e^{-a} \cdot e^{a} = a}$.

Datorită faptului că media estea egală cu **a**, legea lui Poisson apare deseori sub forma: $\psi_n(k) = e^{-\mu}\frac{\mu^k}{k!}$, fără a se uita că $\mu = n.p$;

2. $D^2(X) = \sigma^2 = \mu_2 = a$. Deci, media şi dispersia sînt egale, adică legea Poisson nu depinde decît de un singur parametru, şi anume, μ;

3. $\sigma = \sqrt{n.p} = \sqrt{a}$;

4. **Mo** este egală cu partea întreagă a lui **a**,, $a - 1 < Mo < a$;

5. $CV = \frac{\sigma}{\mu} = \frac{\sqrt{\mu}}{\mu} = \frac{1}{\sqrt{\mu}} = \frac{1}{\sqrt{a}}$;

6. Conform principiului probabilităţii totale, funcţia de repartiţie este:

$F_n(k) = \sum_{i=0}^{k} \frac{a^i e^{-a}}{i!}$. Valorile funcţiei de distribuţie $\psi_n(k)$ şi ale funcţiei de repartiţie $F_n(k)$ sint tabelate;

7. Graficul funcţiei de repartiţie este asimetric, asimetria fiind dictată de valoarea mediei μ, mai ales, pentru valorile mici ale mediei. Prin creşterea lui **n**, vîrful curbei se deplasează spre dreapta şi asimetria curbei scade;

8. Relaţia de recurenţă permite, pe de o parte, determinarea valorilor funcţiei de distribuţie care nu sînt tabelate, iar, de pe alta, excluderea calculării de mai multe ori a aceloraşi valori care intervin în expresia funcţiei de repartiţie:

$$\psi_n(k+1) = \frac{a}{k+1} \cdot \psi_n(k);$$

9. Ca domenii de aplicabilitate a repartiţiei Poisson se enumeră: controlul statistic în industria textilă, biologie şi meteorologie. Pentru **n** suficient de mare şi **p** foarte mic, aşa cum s-a mai spus, repartiţia Poisson poate fi utilizată în locul celei binomiale.

Aplicaţie. Dintr-o urnă se extrage cîte o singură bilă, fără să se pună înapoi. Probabilitatea extragerii unei bile negre este **p = 0.01**. Se formează grupe de cîte **60** extrageri. Numărul de bile negre din diferite grupe oscilează, deci este o variabilă aleatoare. Se cere probabilitatea ca în urma a **60** extrageri să se obţină **3** bile negre?

Se determină: $\mathbf{a = n.p = 60.\ 0.01 = 0.6}$, iar $\mathbf{k = 3}$. Rezultă că:

$$\psi_{60}(3) = (0.6^3 . e^{-0.6})/3! = 0.020.$$

Pentru variabile aleatoare continue, se prezintă următoarele repartiţii teoretice: **normală**, **standard**, **log-normală**, χ^2, **Student** şi **Fisher-Snedecor**. Avînd în vedere că majoritatea lucrărilor de statistică conţin în partea de anexe valorile funcţiei de repartiţie normală standard şi ale quantilelor celorlalte distribuţii continue, din lipsa de spaţiu vom renunţa la prezentarea lor.

7.3. Repartiţia normală

Majoritatea repartiţiilor experimentale pot avea ca model teoretic, cu respectarea unui număr minim de condiţii, repartiţia normală, chiar dacă, de multe ori, gradul de aproximare este scăzut, rezultatele pot fi utilizate cu caracter orientativ. În plus, repartiţia normală poate fi folosită în determinarea unor mărimi statistice care rezultă din studierea unor eşantioane independente raportate la scala interval sau de reper, în vederea desprinderii unor concluzii despre populaţiile din care s-au extras eşantioanele.

Pentru prima dată, repartiţia normală a fost studiată de către Abraham de Moivre în secolul al XVIII-lea. Independent unul de altul, la începutul secolului al XIX-lea, Gauss, studiind distribuţia erorilor întîmplătoare, şi Laplace au descoperit repartiţia cunoscută sub denumirea de repartiţia Gauss-Laplace. Lambert Quetelet (1796 - 1874) a formulat ipoteza ca măsurătorile biologice urmează o repartiţie Gauss, efectuînd o serie de măsurări antropologice. Din acest motiv, repartiţia gaussiană este recunoscută şi sub numele de reparatiţia normală.

Fie **X** o variabilă aleatoare continuă.

Definiţia 7.4.1. Se spune că **X** are repartiţie normală, dacă funcţia ei de repartiţie este dată de:

F(X) = P(X< x) = $\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt$, $x \in \Re$, $\mu \in \Re$, $\sigma > 0$, unde μ şi σ^2 sînt parametrii funcţiei de repartiţie.

Se notează funcţia de repartiţie normală cu $N(\mu; \sigma^2)$, iar v.a **X** repartizată normal cu parametrii μ şi σ^2 cu $X \sim N(\mu; \sigma^2)$.

Densitatea de probabilitate este funcţia:

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

Proprietăţi.

1. $M(X) = Me(X) = Mo(X) = \mu$. Valoarea mediei se obţine rezolvînd integrala: $M(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$. Se arată că prima integrală este zero, iar cea de a doua se rezolvă printr-o schimbare de variabilă: $u = \frac{x-\mu}{\sigma}$, unde $du = \frac{dx}{\sigma}$:

$$M(X) = \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du = \frac{\mu}{\sqrt{2\pi}}.\sqrt{2\pi} = \mu;$$

2. $D^2(X) = \sigma^2$. Valoarea dispersiei rezultă din rezolvarea integralei: $D^2(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 . e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$;

3. $\mu_{2r+1} = 0$ şi $\mu_{2r} = \frac{(2r)!}{2^r r!}\sigma^{2r}$. În particular, $\mu_4 = 3\sigma^4$.

4. $\beta_1 = 0$, $\beta_2 = 3$ şi $E = 0$. Se reamintesc semnificaţiile notaţiilor de mai sus, în ordine: media aritmetică, mediana, moda, dispersia, momentele centrate de ordin **r** (pentru **r** impar şi respectiv par), coeficienţii de asimetrie, de boltire şi respectiv de exces. Se observă că parametrii repartiţiei normale, μ şi σ, au semnificaţia unor valori tipice.

Graficul funcţiei densitate de probabilitatea **f** se numeşte **curba normală** (sau, **curba lui Laplace**) sau **clopotul lui Gauss** şi respectă următoarele proprietăţi:

- este simetric faţă de dreapta $x = \mu$, iar ramurile graficului tind asimptotic la axa absciselor;
- punctul maxim este $\left(\mu, \dfrac{1}{\sigma\sqrt{2\pi}}\right)$;
- repartiţia normală este unimodală;
- în vecinătatea punctului de maxim, convexitatea curbei este spre exterior;
- punctele de abscisă $\mu - \sigma$ şi $\mu + \sigma$ sînt puncte de inflexiune (adică, se schimbă convexitatea curbei);
- modificarea parametrului μ, păstrînd σ constant, produce translaţia curbei, fără modificarea formei;
- modificarea parametrului σ, păstrînd μ constant, produce modificarea formei curbei, o ascuţire a curbei, atunci cînd σ se micşorează sau o aplatizare a curbei, atunci cînd σ creşte;
- valorile măsurate ale caracteristicii studiate tind să se grupeze simetric în jurul mediei (valorii centrale), dînd naştere unui grafic de distribuţie în formă de clopot.

Repartiţie normală cu parametrii $\mu = 1.5$ şi $\sigma = 1$ este reprezentată în figura de mai jos:

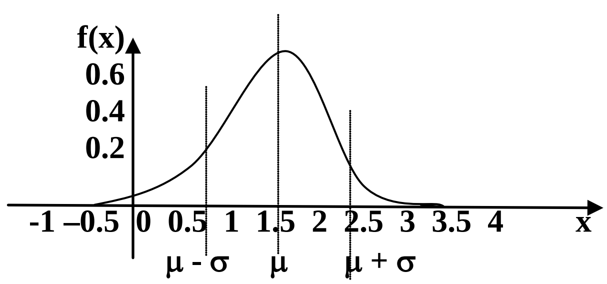

Figura 7.3.1. Clopotul lui Gauss cu $\mu = 1.5$ şi $\sigma = 1$

Cîteva valori remarcabile.

$$P\left(\mu - \frac{2}{3}\sigma < x < \mu + \frac{2}{3}\sigma\right) = 0.50;$$
$$P(\mu - \sigma < x < \mu + \sigma) = 0.68;$$
$$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.95;$$
$$P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.997.$$

Proprietăţi.

1. Dacă $X_1, X_2, \ldots, X_k$ sînt v.a.independente repartizate normal, atunci suma lor $X = X_1 + X_2 + \ldots + X_k$ este o v.a repartizată normal cu parametrii $\mu = \sum_{i=1}^{k} \mu_i$ şi $\sigma^2 = \sum_{i=1}^{k} \sigma_i^2$ unde cu μ_i şi σ_i^2, $i = 1, 2, \ldots, k$ s-au notat mediile şi dispersiile v.a. $X_1, X_2, \ldots, X_k$.

2. Dacă $\mathbf{X_1}$, $\mathbf{X_2}$, **…** , $\mathbf{X_k}$ sînt v.a.independente repartizate normal mediile şi dispersiile $\boldsymbol{\mu_i}$ şi respectiv $\boldsymbol{\sigma_i^2}$, **i = 1, 2, …, k**, atunci orice combinaţie liniară

$\mathbf{X = c_1X_1 + c_2X_2 + \ldots + c_kX_k}$ este o v.a repartizată normal cu parametrii: $\boldsymbol{\mu} = \sum_{i=1}^{k} c_i \mu_i$ şi

$\boldsymbol{\sigma^2} = \sum_{i=1}^{k} c_i^2 \sigma_i^2$, unde $\mathbf{c_i}$**, i = 1, 2, …, k** sînt constante oarecare.