

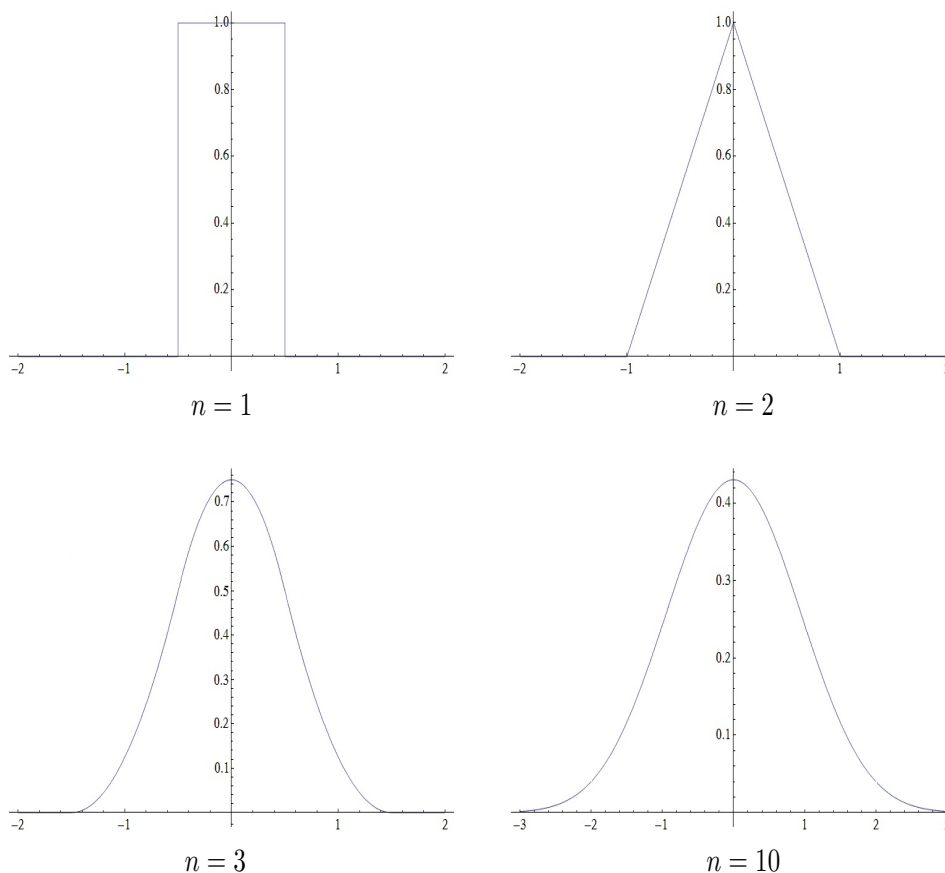
Curs 9

Teorema limită centrală

9.1 Teorema limită centrală. Enunț

Teorema Limită Centrală (TLC) este una dintre cele mai importante teoreme din teoria probabilităților. Intuitiv, teorema afirmă că suma unui număr mare de v. a. independente, identic distribuite, standardizate poate fi aproximată de o distribuție normală. Rezultă importanța acestei distribuții, deși densitatea sa de probabilitate are o formă care pare complicată. Teorema limită centrală mai poartă denumirea de miracolul lui Gauss.

Pentru a realiza acest lucru, să considerăm densitatea de probabilitate a sumei a n variabile independente, de tipul $\text{Unif}[-0.5, 0.5]$. Obținem următoarele:



Dacă vom compara aceste grafice cu graficul unei densități de probabilitate a unei v.a.

normal distribuite, vom observa asemănarea remarcabilă chiar pentru n destul de mic.

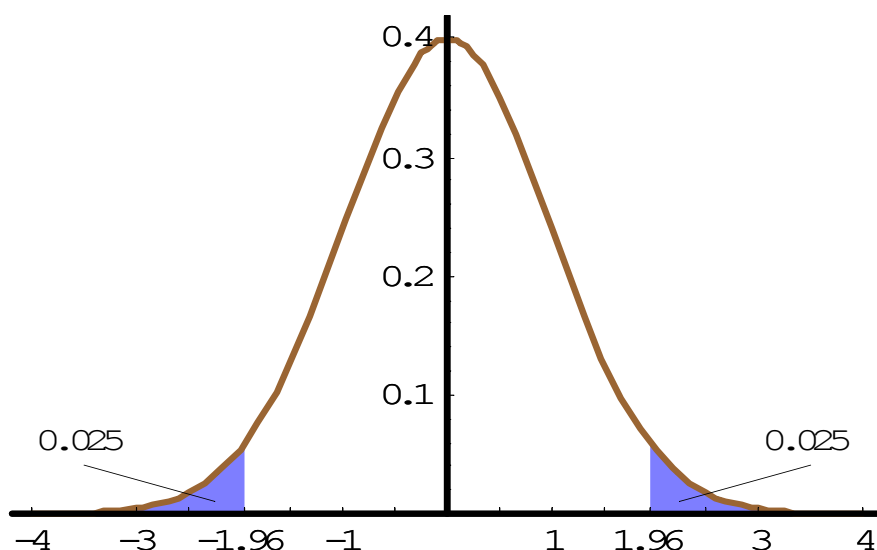


Figura 9.1: Distribuția normală standard

Acest lucru ne conduce intuitiv la ideea că suma variabilelor aleatoare se comportă, într-un anumit sens, normal. Exact Teorema Limită Centrală formalizează acest lucru.

Fie $X_1, X_2, \dots, X_n, \dots$ un șir de v. a. independente și identic distribuite cu $M[X_i] = m$ și $D[X_i] = \sigma^2, i \geq 1$. Fie $S_n = X_1 + X_2 + \dots + X_n$ și $M_n = \frac{1}{n}S_n$. Teorema limită centrală dă informații asupra v. a.

$$Z_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{n(M_n - m)}{n\frac{\sigma}{\sqrt{n}}} = \frac{M_n - m}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{M_n - m}{\sigma}.$$

Deoarece $M[M_n] = m, D[M_n] = \frac{\sigma^2}{n}$, deducem că $M[Z_n] = 0$ și $D[Z_n] = 1$. Această v. a. se numește standardizată, deci șirul Z_n este versiunea standardizată a șirului M_n .

Înainte de a demonstra Teorema Limită Centrală, avem nevoie de următoarele preliminare, ce rezultă din dezvoltările Taylor ale funcției exponențiale:

1. Dacă $u \geq 0$, atunci

$$0 \leq e^{-u} - 1 + u \leq \frac{u^2}{2}. \quad (9.1)$$

2. Dacă $t \in \mathbb{R}$, atunci

$$|e^{it} - 1 - it| \leq \frac{|t|^2}{2} \text{ și } \left| e^{it} - 1 - it - \frac{(it)^2}{2} \right| \leq \frac{|t|^3}{6}. \quad (9.2)$$

De asemenea, vom folosi următorul rezultat:

Teorema 9.1.1 (Teorema convergenței dominate) Fie (f_n) un șir de funcții, convergente punctual la o funcție f , f_n continue cu excepția unui număr finit de puncte și dominate de o funcție integrabilă g :

$$|f_n(x)| \leq g(x), \quad \int_{-\infty}^{\infty} g(x)dx < \infty.$$

Atunci f este integrabilă și

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x)dx = \int_{-\infty}^{\infty} f(x)dx$$

Înainte de a formula TLC, avem nevoie de definiția convergenței în distribuție.

Definiția 9.1.2 Spunem că șirul (X_n) de v.a. **converge în distribuție** la v.a. X și notăm $X_n \xrightarrow{d} X$ dacă

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x),$$

pentru orice punct $x \in \mathbb{R}$ de continuitate a lui F_X .

Formulăm acum un alt rezultat util, ce permite simplificarea demonstrației TLC.

Teorema 9.1.3 (Teorema de continuitate a lui Lévy) Fie (X_n) un șir de v.a. astfel încât șirul funcțiilor caracteristice corespunzătoare φ_{X_n} converge punctual la o funcție φ . Atunci

$$X_n \xrightarrow{d} X \Leftrightarrow \varphi = \varphi_X.$$

Teorema 9.1.4 (Teorema Limită Centrală) Fie (X_n) un șir de v.a. independente, identic distribuite, standardizate ($M[X_i] = 0$ și $D[X_i] = 1$). Atunci, pentru orice $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n}{\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du,$$

adică $\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{d} Z$, unde $Z \in N[0, 1]$.

Demonstrație. Notăm cu $S_n = X_1 + \dots + X_n$. Fie φ_{X_k} funcția caracteristică a lui X_k , care este aceeași pentru fiecare k (i.i.d.), deci o putem nota cu φ . Atunci, pentru orice $t \in \mathbb{R}$,

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = M\left[e^{it\frac{S_n}{\sqrt{n}}}\right] = \prod_{k=1}^n M\left[e^{it\frac{X_k}{\sqrt{n}}}\right] = \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Rămâne de arătat, folosind Teorema de continuitate a lui Levy, că

$$\lim_{n \rightarrow \infty} \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{-\frac{t^2}{2}},$$

deoarece am arătat mai sus că funcția caracteristică a v.a. normale standard este de tipul $e^{-\frac{t^2}{2}}$.

Dacă $t = 0$, nu avem nimic de demonstrat. Presupunem $t \neq 0$. Avem

$$\begin{aligned} \left| \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n - e^{-\frac{t^2}{2}} \right| &= \left| \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n - \left[e^{-\frac{t^2}{2n}}\right]^n \right| \\ &\leq n \left| \varphi\left(\frac{t}{\sqrt{n}}\right) - e^{-\frac{t^2}{2n}} \right|, \end{aligned} \quad (9.3)$$

deoarece $\left|\varphi\left(\frac{t}{\sqrt{n}}\right)\right| \leq 1$ și $0 \leq e^{-\frac{t^2}{2n}} \leq 1$.

Urmează

$$n \left| \varphi\left(\frac{t}{\sqrt{n}}\right) - e^{-\frac{t^2}{2n}} \right| \leq n \left| \varphi\left(\frac{t}{\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right) \right| + n \left| \left(1 - \frac{t^2}{2n}\right) - e^{-\frac{t^2}{2n}} \right|.$$

Folosind (9.1) pentru $u = \frac{t^2}{2n}$, obținem

$$n \left| \left(1 - \frac{t^2}{2n}\right) - e^{-\frac{t^2}{2n}} \right| \leq n \frac{t^4}{8n^2} = \frac{t^4}{8n} \rightarrow 0 \text{ pentru } n \rightarrow \infty. \quad (9.4)$$

Pentru primul modul,

$$\begin{aligned} n \left| \varphi \left(\frac{t}{\sqrt{n}} \right) - \left(1 - \frac{t^2}{2n} \right) \right| &= n \left| M \left[e^{i \frac{t}{\sqrt{n}} X} - \left(1 + i \frac{t}{\sqrt{n}} X + i^2 \frac{t^2}{2n} X^2 \right) \right] \right| \\ &\leq n \cdot M \left[\left| e^{i \frac{t}{\sqrt{n}} X} - \left(1 + i \frac{t}{\sqrt{n}} X + i^2 \frac{t^2}{2n} X^2 \right) \right| \right], \end{aligned} \quad (9.5)$$

deoarece $M[X] = 0$ și $M[X^2] = D[X] = 1$.

Pe de o parte, folosind prima relație din (9.2), obținem

$$\begin{aligned} \left| e^{i \frac{t}{\sqrt{n}} X} - \left(1 + i \frac{t}{\sqrt{n}} X + i^2 \frac{t^2}{2n} X^2 \right) \right| &\leq \left| e^{i \frac{t}{\sqrt{n}} X} - \left(1 + i \frac{t}{\sqrt{n}} X \right) \right| + \frac{t^2}{2n} X^2 \\ &\leq \frac{t^2}{2n} X^2 + \frac{t^2}{2n} X^2 = \frac{t^2}{n} X^2. \end{aligned}$$

Pe de altă parte, folosind a doua relație din (9.2), obținem

$$\left| e^{i \frac{t}{\sqrt{n}} X} - \left(1 + i \frac{t}{\sqrt{n}} X + i^2 \frac{t^2}{2n} X^2 \right) \right| \leq \frac{|t|^3 |X|^3}{6n^{3/2}}.$$

Pentru orice $\delta > 0$ și $n \in \mathbb{N}^*$, definim

$$A := A(\delta, n) := \{|X| > \delta\sqrt{n}\}$$

și funcția caracteristică a mulțimii A ,

$$I_A(x) = \begin{cases} 1, & \text{dacă } x \in A \\ 0, & \text{în rest.} \end{cases}$$

Atunci

$$\left| e^{i \frac{t}{\sqrt{n}} X} - \left(1 + i \frac{t}{\sqrt{n}} X + i^2 \frac{t^2}{2n} X^2 \right) \right| \leq \frac{t^2}{n} X^2 I_A + \frac{|t|^3 |X|^3}{6n^{3/2}} I_{A^c}.$$

Rezultă

$$\begin{aligned} n \cdot M \left[\left| e^{i \frac{t}{\sqrt{n}} X} - \left(1 + i \frac{t}{\sqrt{n}} X + i^2 \frac{t^2}{2n} X^2 \right) \right| \right] &\leq n \frac{t^2}{n} M[X^2 I_A] + n \frac{|t|^3}{6n^{3/2}} M[|X|^3 I_{A^c}] \\ &= t^2 M[X^2 I_A] + \frac{|t|^3}{6n^{1/2}} M[|X|^3 I_{A^c}]. \end{aligned} \quad (9.6)$$

Cum $|X| \leq \delta\sqrt{n}$ pe A^c , rezultă

$$M[|X| I_{A^c}] \leq \delta\sqrt{n} \underbrace{\int_{-\infty}^{\infty} I_{A^c} \cdot f_X(x) dx}_{\leq 1},$$

de unde

$$M[|X|^3 I_{A^c}] \leq M[X^2] \cdot M[|X| I_{A^c}] \leq \delta\sqrt{n}.$$

Deci,

$$\frac{|t|^3}{6n^{1/2}} M[|X|^3 I_{A^c}] \leq \frac{|t|^3 \delta}{6}.$$

De asemenea,

$$M[X^2 I_A] = 1 - M[X^2 I_{\{|X| \leq \delta\sqrt{n}\}}].$$

Cum șirul de funcții $f_n := X^2 I_{\{|X| \leq \delta \sqrt{n}\}} f_X$ este crescător, cu limita punctuală $f = X^2 f_X$, care este integrabilă ($M[X^2] = 1$), rezultă aplicând Teorema convergenței dominate că $\lim_{n \rightarrow \infty} M[X^2 I_{\{|X| \leq \delta \sqrt{n}\}}] = M[X^2] = 1$.

Fie acum $\varepsilon > 0$. Fie $\delta > 0$ astfel încât $\frac{|t|^3 \delta}{6} < \frac{\varepsilon}{4}$. Alegem de asemenea $n_\varepsilon \in \mathbb{N}$ astfel încât, pentru orice $n \geq n_\varepsilon$, să avem $\frac{t^4}{8n} < \frac{\varepsilon}{4}$ și $t^2 M[X^2 I_A] < \frac{\varepsilon}{2}$. Va rezulta, combinând relațiile (9.3)-(9.6), că pentru orice $\varepsilon > 0$, există $n_\varepsilon \in \mathbb{N}$ astfel încât, pentru orice $n \geq n_\varepsilon$,

$$\left| \left[\varphi \left(\frac{t}{\sqrt{n}} \right) \right]^n - e^{-\frac{t^2}{2}} \right| < \varepsilon,$$

ceea ce era de demonstrat. □

Alte variante ale teoremei limită centrală sunt următoarele:

Teorema 9.1.5 (TLC1) Fie $(X_n)_{n \in \mathbb{N}}$ un șir de v. a. independente și identic distribuite (i.i.d.) cu $M[X_i] = m$ și $D[X_i] = \sigma^2, i \geq 1$. Atunci pentru orice numere reale $a < b$, avem

$$\lim_{n \rightarrow \infty} P(\{a \leq Z_n \leq b\}) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx. \quad (9.7)$$

Teorema 9.1.6 (TLC2) Fie $(X_n)_{n \in \mathbb{N}}$ un șir de v. a. independente distribuite cu $M[X_i] = m$ și $D[X_i] = \sigma_i^2, i \geq 1$ și $\lim_{n \rightarrow \infty} \sigma_n^2 = \sigma^2$. Atunci pentru orice numere reale $a < b$, avem

$$\lim_{n \rightarrow \infty} P(\{a \leq Z_n \leq b\}) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

9.2 Aplicații ale Teoremei limită centrală

Exercițiul 1 Aplicație la sondaje de opinie. Presupunem că se realizează experimente de tip Bernoulli în care un eveniment A se produce cu probabilitatea p . Notăm cu X_k variabila aleatoare care ia valoarea 1 dacă la experiența cu numărul de ordine k se produce evenimentul A și 0 dacă nu se produce A . Variabilele aleatoare $(X_k)_{k \geq 1}$ sunt independente. Atunci $S_n = X_1 + X_2 + \dots + X_n$ reprezintă numărul total de apariții ale lui A , deci numărul de succese ale lui A în urma efectuării a n experiențe. S_n este o v. a. repartizată binomial, $S_n \in Bi(n, p)$. V. a. $(X_n)_{n \in \mathbb{N}}$ au aceeași repartiție (binomială).

Notăm cu $Z_n = \frac{S_n - np}{\sqrt{npq}}$. Conform teoremei limită centrală rezultă că pentru orice $a < b$ și n suficient de mare,

$$P(\{a \leq Z_n \leq b\}) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx = \Phi(b) - \Phi(a).$$

Pentru orice $\alpha < \beta$, avem $\alpha \leq S_n \leq \beta$ dacă și numai dacă $\frac{\alpha - np}{\sqrt{npq}} \leq Z_n \leq \frac{\beta - np}{\sqrt{npq}}$.

Notând $\frac{\alpha - np}{\sqrt{npq}} = a, \frac{\beta - np}{\sqrt{npq}} = b$, rezultă că $P(\{\alpha \leq S_n \leq \beta\}) = \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right)$

și deci

$$P(np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}) \cong \Phi(b) - \Phi(a).$$

În particular, pentru $a = -b$ ($b > 0$) rezultă formula

$$P(\{np - b\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}\}) \cong 2 \cdot \Phi(b) - 1, \quad (9.8)$$

pentru $n \gg 1$ (unii statisticieni recomandă $npq \geq 10$).

Această formulă este utilizată în sondaje astfel: considerăm o populație statistică umană căreia îi cerem opinia într-o anumită chestiune: ce echipă de fotbal, ce partid, ce televiziune etc. preferă. Nu toată lumea poate fi consultată și atunci se realizează un sondaj pe eșantioane restrânse, alese cu obiectivitate. Să presupunem că se consultă n persoane și notăm cu S_n numărul de persoane care se pronunță pentru (succes); se determină parametrul p ca fiind frecvența de succes. Pentru $b = 2.17$ avem $\Phi(b) = 0.985$ deci $2 \cdot \Phi(b) - 1 = 0.97$ și conform (9.8), se realizează cu eroare sub 3% evenimentul $S_n \in [np - 2.17\sqrt{npq} \leq S_n \leq np + 2.17\sqrt{npq}]$; apoi pentru $b = 1.96$ avem $2 \cdot \Phi(b) - 1 = 0.952$ și conform (9.8), se realizează cu eroare sub 5% evenimentul

$$S_n \in [np - 1.96\sqrt{npq} \leq S_n \leq np + 1.96\sqrt{npq}].$$

Exercițiul 2 Dintr-un sondaj realizat într-un oraș a rezultat că dintr-un eșantion de 1000 votanți 600 ar vota cu partidul X. Cu o eroare de sub 3% să se estimeze câți dintre cei 1,2 milioane de votanți ar vota pentru X.

Soluție. Avem $p = \frac{600}{1000} = 0.6$ și $q = 0.4$, $n = 1200000$ și luăm $b = 2.17$, deci numărul cerut este cuprins între $np - 2.17\sqrt{npq}$ și $np + 2.17\sqrt{npq}$ deci între $1200000 \cdot 0.6 - 2.17\sqrt{1200000 \cdot 0.6 \cdot 0.4} = 718840$ și $1200000 \cdot 0.6 + 2.17\sqrt{1200000 \cdot 0.6 \cdot 0.4} = 721160$.

9.2.1 Aproximarea legii binomiale printr-o lege normală

Fie X o v. a. discretă repartizată binomial cu parametrii p și n . (deci $M[X] = np$, $D[X] = npq$ și $k \in N, k \leq n$). Să se calculeze $P(\{X = k\})$ și $P(\{X \leq k\})$ folosind teorema limită centrală (pentru n suficient de mare și pq nu foarte mic).

Conform teoremei limită centrală $X \in N(np, npq)$ și deci

$$P(\{X = k\}) = P(\{X \in [k - \frac{1}{2}, k + \frac{1}{2}]\}) = \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{npq}}\right).$$

Adăugarea lui 0.5 la k se numește corecție prin continuitate. A fost pentru îmbunătățirea aproximației.

$$\text{La fel } P(\{X \leq k\}) = P(\{X \leq k + \frac{1}{2}\}) = \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{npq}}\right).$$

Exercițiul 3 Se aruncă o monedă și probabilitatea de a obține banul este 0.6. Se aruncă moneda de 1000 ori. Care este probabilitatea de a obține banul de 650 de ori?

Soluție. Fie X v. a. care ia ca valori numărul de apariții ale banului în cele 1000 de aruncări. Evident $X \in \text{Binomial}[1000; 0.6]$. Conform teoremei limită centrală $X \in N(600, 240)$.

$$P(\{X = 650\}) = P(\{X \in [650 - \frac{1}{2}, 650 + \frac{1}{2}]\}) = \Phi\left(\frac{650 + \frac{1}{2} - 600}{\sqrt{240}}\right) - \Phi\left(\frac{650 - \frac{1}{2} - 600}{\sqrt{240}}\right) = \Phi(3.25) - \Phi(3.19) = 0.999423 - 0.999289 = 0.000134. \blacktriangledown$$

O v. a. distribuită binomial se aproximează cu o v. a. distribuită normal dacă $np > 5, nq > 5$.

9.2.2 Aproximarea legii Poisson printr-o lege normală

Fie X o v. a. discretă repartizată Poisson cu parametru λ . Atunci $M[X] = \lambda, D[X] = \lambda$. Conform teoremei limită centrală $X \in N(\lambda, \lambda)$. Să se calculeze $P(\{X = k\})$ și $P(\{X \leq k\})$ pentru k număr natural.

$$P(\{X = k\}) = P(\{X \in [k - \frac{1}{2}, k + \frac{1}{2}]\}) = \Phi\left(\frac{k + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{k - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right).$$

La fel $P(\{X \leq k\}) = P(\{X \leq k + \frac{1}{2}\}) = \Phi\left(\frac{k + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$.

Aproximarea este bună dacă $\lambda > 5$.

Exercițiul 4 Statistica arată că la o unitate de asigurări se primesc în medie 300 de reclamații pe an. Fie X numărul de reclamații pe an, presupus repartizat Poisson. Să se determine probabilitatea ca să primească cel puțin 351 de reclamații pe an.

Soluție. $P(\{X \geq 351\}) = 1 - P(\{X \leq 351\}) = 1 - \Phi\left(\frac{351.5 - 300}{\sqrt{300}}\right) = 1 - \Phi(2.9734) = 1 - 0.99853 = 0.00147$. ▼