Лекція 3

Повторні незалежні випробування. Формула Бернуллі. Граничні теореми

Розглянемо експеримент, що складається з незалежних випробувань, в кожному з яких деяка подія може настати або не настати зі сталою ймовірністю. Ця ймовірність не зміниться і в тому випадку, коли буде відомо, чи наставала вона в результаті попередніх випробувань, чи ні.

4.1. Схема незалежних випробувань. Формула Бернуллі

Нехай проводиться яке завгодно скінчене число n однакових випробувань, в кожному з яких можливими є два несумісні результати: подія A настане з певною ймовірністю p або не настане (подія $\overline{A}$) з ймовірністю $1-p$. Такі випробування називають *незалежними повторними випробуваннями відносно події* A.

Серію незалежних повторних випробувань з одним із можливих результатів A або $\overline{A}$, у кожному з яких подія A має сталу ймовірність появи $P(A)=p$, називають *послідовністю незалежних випробувань або схемою Бернуллі*.

Поставимо задачу: обчислити ймовірність $P_n(m)$ того, що в результаті проведення n незалежних випробувань подія A настане рівно m разів, якщо в кожній із спроб вона настає із сталою ймовірністю $P(A)=p$ або не настає з ймовірністю $P(\overline{A})=1-P(A)=1-p=q$. Зауважимо, що тут не вимагається, щоб подія A відбулася m разів в певній послідовності.

Для розв´язання поставленої задачі можна безпосередньо використати теореми додавання і множення ймовірностей: ця подія є сумою несумісних подій, кожна з яких дорівнює добутку m подій A та $n-m$ подій $\overline{A}$. Ймовірність кожного з можливих доданків за теоремою множення ймовірностей незалежних подій дорівнює $p^m \cdot q^{n-m}$. Таких доданків стільки, скільки є всеможливих варіантів розмістити m подій A по n місцях. Їх кількість дорівнює кількості комбінацій C_n^m з n по m. Цьому можна дати чітке формулювання:

Формула Бернуллі. *Якщо ймовірність p настання події A в кожному випробуванні стала і залишається незмінною, коли стають відомими результати попередніх випробувань, то ймовірність того, що в n випробуваннях подія A настане рівно m разів дорівнює*

$$P_n(m) = C_n^m \cdot p^m \cdot q^{n-m} = \frac{n!}{m!(n-m)!} p^m \cdot (1-p)^{n-m}. \qquad (4.1)$$

Приклад 4.1. Гральний кубик підкидають п'ять разів. Яка ймовірність того, що при цьому тричі випаде 6 очок?

Розв'язання. Нехай подія A – при одному киданні випаде 6 очок; ймовірність цієї події $P(A) = p = \frac{1}{6}$, відповідно $q = 1 - p = \frac{5}{6}$. Згідно запропонованих вище позначень, у цій задачі $n = 5$, $m = 3$. Отже, за формулою (3.1) отримаємо

$$P_5(3) = C_5^3 \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^2 = \frac{5!}{3!\cdot 2!} \cdot \frac{25}{6^5} = \frac{250}{7776} \approx 0{,}032.$$

(Цей результат потрібно трактувати так: якщо дослід із п'ятьма підкиданнями кубика проводити багато разів, то в середньому в 250 випадках із 7776 грань з 6 очками випаде рівно три рази).

Зауваження. Ймовірності $P_n(m)$ називаються *біномними*, оскільки вони мають відношення до формули бінома Ньютона:

$$(q+p)^n = q^n + C_n^1 \cdot p^1 \cdot q^{n-1} + C_n^2\ p^2 \cdot q^{n-2} + \ldots + C_n^m \cdot p^m \cdot q^{n-m} + \ldots + C_n^{n-1} \cdot\ p^{n-1} \cdot q^1 + p^n,$$

або

$$(q+p)^n = P_n(0) + P_n(1) + P_n(2) + \ldots + P_n(m) + \ldots + P_n(n-1) + P_n(n).$$

Остання формула підтверджує той факт, що події: у серії з n незалежних випробувань подія настала 0 разів (ймовірність дорівнює $P_n(0)$), 1 раз (ймовірність $P_n(1)$), 2 рази (ймовірність $P_n(2)$), …, всі n разів (ймовірність $P_n(n)$), утворюють повну групу попарно несумісних подій. Справді, оскільки $q + p = 1$, то й $(q+p)^n = 1$. Звідки

$$P_n(0) + P_n(1) + P_n(2) + \ldots + P_n(n-1) + P_n(n) = 1.$$

Приклад 4.2. В гаражі деякого підприємства є 10 автомашин. Для нормальної роботи підприємства необхідно мати на лінії не менше 7 автомашин. Ймовірність невиходу на лінію для кожної автомашини дорівнює 0,1. Знайти ймовірність нормальної роботи підприємства в найближчий день.

Розв'язання. Підприємство буде нормально працювати (подія B), якщо на лінію вийдуть 7, 8, 9 або всі 10 автомашин (події A_1, A_2, A_3, A_4). Ймовірність виходу на лінію для кожної автомашини дорівнює

$p = 1 - 0{,}1 = 0{,}9$ і невиходу $q = 0{,}1$. Тоді згідно з теоремою про ймовірність суми несумісних подій

$$P(B) = P(A_1 + A_2 + A_3 + A_4) = P(A_1) + P(A_2) + P(A_3) + P(A_4) =$$
$$= P_{10}(7) + P_{10}(8) + P_{10}(9) + P_{10}(10) = 0{,}9872.$$

Зауваження. У випадку, коли подія B складається із більшої кількості доданків, ніж подія $\overline{B}$, доцільно спочатку знаходити ймовірність події $\overline{B}$, а тоді $P(B) = 1 - P(\overline{B})$. Зокрема, якщо потрібно знайти ймовірність того, що у n незалежних випробуваннях подія A настане хоча б один раз (подія B), то протилежною до неї є подія $\overline{B}$ – у n незалежних випробуваннях подія A не настане жодного разу. Тоді ймовірність $P(B)$ доцільно шукати за формулою $P(B) = 1 - P_n(0) = 1 - q^n$.

4.2. Найімовірніше число появ події

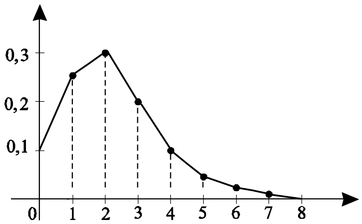
Тепер звернемо увагу на деякі елементарні факти, які стосуються поведінки ймовірності $P_n(m)$, як функції аргумента m при сталому значенні кількості n незалежних випробувань.

Нехай ймовірність того, що студенту потрібно взяти в бібліотеці книгу з математичної дисципліни дорівнює 0,25. Серед восьми перших відвідувачів бібліотеки якогось дня може не бути жодного, який замовив би книгу з математики, може бути один, два, три, … , всі вісім. Використовуючи формулу Бернуллі при $n = 8$, $p = 0{,}25$, $q = 0{,}75$, $m = 0,1,2,\ldots,8$ знайдемо ймовірність кожної з цих подій:

$$P_8(0) \approx 0{,}1001 \qquad P_8(5) \approx 0{,}0231$$
$$P_8(1) \approx 0{,}2670 \qquad P_8(6) \approx 0{,}0038$$
$$P_8(2) \approx 0{,}3115 \qquad P_8(7) \approx 0{,}0004$$
$$P_8(3) \approx 0{,}2076 \qquad P_8(8) \approx 0{,}0000$$
$$P_8(4) \approx 0{,}0865$$

Зауважимо, що сума всіх дев'яти ймовірностей дорівнює 1, оскільки відповідні їм події утворюють повну групу несумісних подій. Таке сумування слугує перевірці правильності проведених обчислень.

У прямокутній системі координат вздовж горизонтальної осі відкладемо значення $m = 0, 1, 2, 3, \ldots, 8$, а по вертикальній – відповідні їм значення ймовірностей $P_8(0), P_8(1), \ldots, P_8(8)$. Побудуємо так званий *многокутник розподілу ймовірностей* кількості настань події. Для цього сполучимо ламаною лінією точки з координатами $(m, P_8(m))$.



З рисунка чітко видно, що із зростанням m ймовірність $P_8(m)$ спочатку також росте, а досягнувши найбільшого значення 0,3115 при $m=2$, починає спадати і практично стає нульовою при $m=8$. Найбільша ордината відповідає найбільшій ймовірності, тоді як абсциса точки з найбільшою ординатою вказує на найімовірніше число настань події в серії незалежних випробувань (таке число згодом ми назвемо *модою* для відповідної випадкової величини). Рисунок ілюструє, що найбільш ймовірно рівно 2 студенти з 8, які першими прийшли в бібліотеку певного дня, попросять книгу з математики.

Найімовірнішим числом появ події *A в n незалежних випробуваннях називається число m_0, якому відповідає найбільше значення ймовірності $P_n(m)$.*

Подвійна нерівність, з якої визначається найімовірніше число m_0 появ події, має вигляд:

$$np-q \le m_0 \le np+p\,, \tag{4.2}$$

Приклад 4.3. Статистичне дослідження показує, що близько 85% виборців планує прийти на вибори. Знайти найімовірніше число мешканців будинку, які прийдуть на вибори, якщо у ньому проживає 150 виборців.

Розв'язання. За умовою задачі $p=0{,}85,\ q=0{,}15,\ n=150.$ Із нерівності (3.6) маємо $150\cdot 0{,}85-0{,}15 \le m_0 \le 150\cdot 0{,}85+0{,}85$ або $127{,}35 \le m_0 \le 128{,}35$, звідки $m_0=128.$

Зауважимо, що довжина інтервалу, якому належить m_0 дорівнює $1: (np+p)-(np-q)=p+q=1.$ Тому, якщо межі цього інтервалу – дробові числа, то отримаємо тільки одне значення m_0, якщо ж межі є цілими числами, то отримаємо два значення найімовірнішого числа появ події: $m_0'=np+p$ та $m_0''=np-q$.

Варто відзначити, що у випадку $np\in Z$ (ціле число), то саме цей добуток дорівнює найімовірнішому числу настань події в n незалежних випробуваннях, тобто $m_0=np$. Добуток np називають *середнім числом появ події в n незалежних випробуваннях.*

4.3. Локальна теорема Муавра–Лапласа

Очевидно, що обчислення за формулою Бернуллі будуть надзвичайно громіздкими, якщо n достатньо велике, зокрема із-за великої кількості множників похибка заокруглень буде нагромаджуватись. Тому доцільно замінити її нехай наближеною, але такою формулою, яка спростила б обчислення, але при цьому забезпечувала б достатньо точні значення шуканих ймовірностей.

Теорема 1 (локальна теорема Муавра-Лапласа). *Якщо ймовірність* p $(0 < p < 1)$ *появи події* A *в кожному випробуванні стала, а кількість випробувань достатньо велика, то ймовірність* $P_n(m)$ *того, що в* n *спробах подія* A *настане рівно* m *разів наближено дорівнює* $P_n(m) \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

або

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi(x), \tag{4.3}$$

де $x = \frac{m - np}{\sqrt{npq}}$, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

При великих n формула (4.3) дає достатньо точне значення ймовірностей $P_n(m)$, причому чим більше n, тим точнішими є обчислені значення. Такі формули для обчислення наближених значень називають *асимптотичними.*

Функцію $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ називають *функцією Гаусса*. На рис. 4 зображений графік цієї функції, який носить назву *гауссової кривої* (або *кривої Гаусса).* Функція Гауса табульована і володіє очевидними **властивостями**:

1^0 $\varphi(x)$ – парна, тобто $\varphi(-x) = \varphi(x)$ (змінна входить у парному степені);

2^0 $\varphi(x) > 0$ для будь-яких $x \in R$;

3^0 при $x \in (-\infty; 0)$ $\varphi(x)$ монотонно зростає, досягає найбільшого значення $\frac{1}{\sqrt{2\pi}}$ при $x = 0$ і при $x \in (0; +\infty)$ – монотонно спадає;

4^0 $\lim\limits_{x\to\infty} \varphi(x) = \lim\limits_{x\to\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \lim\limits_{x\to\infty} \frac{1}{\sqrt{2\pi} e^{\frac{x^2}{2}}} = 0$, і вже при $|x| \geq 5$ можна вважати, що $\varphi(x) \approx 0$.

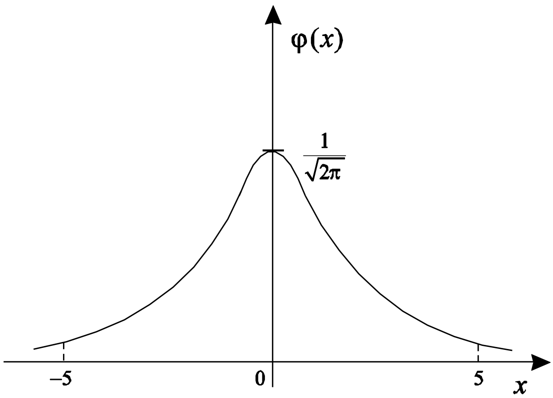


Рис. 4. Функція Гаусса

Приклад 4.4. Яка ймовірність того, що подія A наступить рівно 80 разів в 400 незалежних випробуваннях, якщо ймовірність появи події в кожній спробі $p = 0{,}2$.

Розв'язання. Тут $n = 400$, $m = 80$, $p = 0{,}2$, $q = 0{,}8$. Обчислимо $\sqrt{npq}$ та x: $\sqrt{npq} = \sqrt{400 \cdot 0{,}8 \cdot 0{,}2}$ =8, $x = \dfrac{80 - 400 \cdot 0{,}2}{8} = 0$. За таблицею значень функції $\varphi(x)$ знаходимо $\varphi(0) = 0{,}3989$. Отже, $P_{400}(80) \approx \dfrac{0{,}3989}{8} \approx 0{,}04986.$

Очевидно, що підрахунок ймовірності за формулою Муавра – Лапласа є досить простим. Підрахунок за формулою Бернуллі дає аналогічний наближений результат $P_{400}(80) \approx 0{,}0498$.

Слід зауважити, що значення ймовірності $P_n(m)$, обчислене за формулою (4.3) дає добре наближення ймовірності настання події рівно m разів у n випробуваннях, якщо ймовірність p настання події в одному випробуванні не є близькою ні до 0, ні до 1. У випадках $p \approx 0$ або $p \approx 1$ краще наближення дає інша формула, про яку йтиметься в пункті 3.6.

4.4. Інтегральна теорема Лапласа

В переважній більшості практичних задач досліднику не потрібно знати ймовірність настання події рівно m разів у n випробуваннях (такі події є дуже малоймовірними при достатньо великих n). Проте у практичних задачах часто потрібно знаходити ймовірності $P_n(m_1, n)$ того, що подія настала не менше m_1 разів, або ймовірності $P_n(0, m_2)$ настання події не більше m_2 разів .

Нехай потрібно знайти ймовірність того, що серед 1000 новонароджених є від 400 до 500 дівчаток, якщо ймовірність народження дівчинки дорівнює 0,49. Цю задачу теоретично можна розв’язати, застосовуючи локальну теорему Муавра – Лапласа і теорему про ймовірність суми несумісних подій:

$$P_{1000}(400,500) = P_{1000}(400) + P_{1000}(401) + P_{1000}(402) + \ldots + P_{1000}(500).$$

Всього потрібно було б 101 раз застосувати формулу Муавра – Лапласа, після чого додати отримані числа. Крім громіздкості обчислень, серйозним недоліком такого підходу до розв’язання є накопичення похибок. Значно оптимальніший підхід до розв’язання сформульованої і аналогічних до неї задач пропонує наступна теорема.

Теорема 2 (інтегральна теорема Муавра - Лапласа). *Якщо ймовірність p появи події A в кожній спробі стала і така, що $0 < p < 1$, то при великих n ймовірність $P_n(m_1, m_2)$ того, що подія A настане в n випробуваннях не менше m_1 і не більше m_2 разів, наближено дорівнює*

$$P_n(m_1, m_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{z^2}{2}} dz \tag{4.4}$$

або

$$P_n(m_1, m_2) \approx \Phi(x_2) - \Phi(x_1), \tag{4.5}$$

де $x_1 = \dfrac{m_1 - np}{\sqrt{npq}}$, $x_2 = \dfrac{m_2 - np}{\sqrt{npq}}$; $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-\frac{z^2}{2}} dz$.

Формула Муавра – Лапласа (4.5) дає достатньо точне наближення ймовірності, якщо значення добутку npq досягає декількох сотень. Якщо задача не вимагає високої точності, то допустими є випадки $npq \geq 20$.

Функція $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-\frac{z^2}{2}} dz$ носить назву *функції Лапласа.*

*Ф*ункція Лапласа табульована та має такі властивості**:**

1^0 Функція $\Phi(x)$ є непарною, тобто $\Phi(-x) = -\Phi(x)$.

2^0 Функція $\Phi(x)$ монотонно зростає на R.

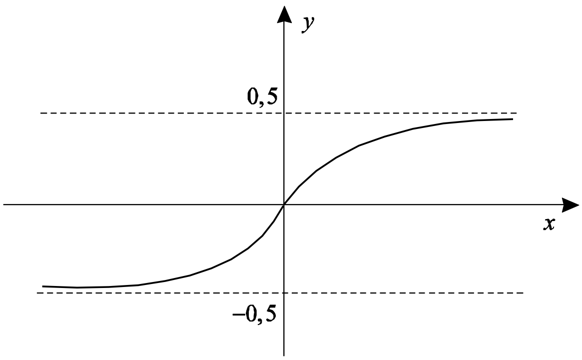


Рис. 5. Функція Лапласа

3[0] $\lim\limits_{x\to\infty} \Phi(x) = 0{,}5$. Цю властивість на практиці застосовують так: оскільки вже при $x = 5$ значення функції Лапласа $\Phi(5) = 0{,}499997$ і $\lim\limits_{x\to\infty} \Phi(x) = 0{,}5$, то вважають, що $\Phi(x) = 0{,}5$ при $x > 5$.

Приклад 4.5. Бухгалтери деякого підприємства в середньому роблять помилки у 15% бухгалтерських звітів. Для перевірки вибирають 200 звітів. Яка ймовірність того, що перевірка виявить від 20 до 30 звітів з помилками?

Розв'язання. Тут $n = 200,\ p = 0{,}15,\ q = 0{,}85,\ m_1 = 20,\ m_2 = 30.$

Обчислимо $\sqrt{npq} = \sqrt{200 \cdot 0{,}15 \cdot 0{,}85} \approx 5{,}05$, $x_1 = \dfrac{20 - 200 \cdot 0{,}15}{5{,}05} \approx -1{,}98$,

$x_2 = \dfrac{30 - 200 \cdot 0{,}15}{5{,}05} = 0$ Отже, за формулою (3.9) маємо

$$P_{200}(20; 30) \approx \Phi(0) - \Phi(-1{,}98) = 0 + 0{,}4761 = 0{,}4761.$$

Розглянемо випадок, коли числа m_1 і m_2 є рівновіддаленими відносно середнього числа np настань події A, тобто існує деяке додатне число r таке, що $m_1 = np - r$, $m_2 = np + r$, тоді

$P(m_1 \le m \le m_2) = P(np - r \le m \le np + r) = P(|m - np| \le r)$,

а формула (3.9) набуде такого вигляду

$$P(|m - np| \le r) = \Phi\left(\frac{np + r - np}{\sqrt{npq}}\right) - \Phi\left(\frac{np - r - np}{\sqrt{npq}}\right) =$$

$$= \Phi\left(\frac{r}{\sqrt{npq}}\right) - \Phi\left(\frac{-r}{\sqrt{npq}}\right) = 2\Phi\left(\frac{r}{\sqrt{npq}}\right). \qquad (4.6)$$

Ця формула означає, що ймовірність відхилення кількості m настань події A в n незалежних випробуваннях від середнього числа np настань події A на деяку додатну величину r дорівнює подвоєному значенню функції Лапласа в точці $\frac{r}{\sqrt{npq}}$.

4.6. Теорема Пуассона

Розглянемо випадок, коли асимптотичні формули не дають достатньої точності: це трапляється, коли ймовірність p настання події в одному незалежному випробуванні є близькою до 0 або 1. Події за умови $p \approx 0$ називають масовими, але рідкісними.

У випадку масових, але рідкісних подій хороше наближення значення ймовірності $P_n(m)$ пропонує теорема Пуассона.

Теорема 3 (теорема Пуассона). *Якщо в серії незалежних випробувань їх кількість* n *велика, а ймовірність* p *настання події* A в *одному випробуванні близька до нуля (звичайно,* $0 < p < 0{,}1$, $npq \le 10$*), то ймовірність* $P_n(m)$ *настання події в* n *незалежних випробуваннях рівно* m *разів наближено дорівнює*

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}, \tag{4.7}$$

де $\lambda = np$.

Ця формула називається *формулою Пуассона*.

Ймовірність, обчислена за формулою (4.7) є достатньо близькою до точного значення ймовірності $P_n(m)$ у випадку $\lambda \le 10$.

Для функції $P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}$ існують таблиці значень при фіксованих значеннях n.

Якщо необхідно за умов теореми Пуассона обчислити ймовірність $P_n(m_1, m_2)$ того, що подія з’явиться не менше m_1 і не більше m_2 разів, то, сумуючи ймовірності для $m_1 \le m \le m_2$, знайдені за формулою Пуассона, отримують другу формулу Пуассона

$$P_n(m_1, m_2) = \sum_{m=m_1}^{m_2} P_n(m) \approx \sum_{m=m_1}^{m_2} \frac{\lambda^m}{m!} e^{-\lambda}. \tag{4.8}$$

Приклад 4.6. Ймовірність виграшу на кожен акційний товар дорівнює 0,001. Яка ймовірність того, що серед 2000 акційних товарів виявиться 4 виграшних?

Розв'язання. За умовою задачі $p = 0{,}001,\ n = 2000,\ m = 4$, звідки $\lambda = np = 2 < 10,$ що забезпечує ефективність застосування формули Пуассона. За формулою (4.7) отримаємо $P_{2000}(4) \approx \frac{2^4}{4!} e^{-2} \approx 0{,}090224.$

4.7. Найпростіший потік подій

Нехай у випадкові моменти часу одна за одною відбуваються деякі однотипні події (на станцію прибувають поїзди, в порт пришвартовуються кораблі, до диспетчера надходять сигнали, у крамницю навідуються покупці, тощо). Таку послідовність подій називають *потоком подій.*

Середню кількість подій, які відбуваються за одиницю часу, називають *інтенсивністю потоку подій* і позначають λ. Потік називають *стаціонарним*, якщо його інтенсивність λ є сталою, тобто середня кількість настань події не залежить від початкового моменту, а залежить лише від довжини часового проміжку. Отже, для стаціонарного потоку подій за одиницю часу відбувається в середньому λ подій.

Якщо ймовірність настання події більше, ніж один раз за досить короткий проміжок часу прямує до нуля, то такий потік називають *ординарним*.

Якщо інтенсивність потоку в наступний проміжок часу не залежить від того, скільки подій настало у попередні проміжки, то таку властивість потоку називають *відсутністю післядії*.

Найпростішим (пуассонівським) називають потік подій, який має властивість стаціонарності, ординарності та відсутністю післядії.

Для найпростішого потоку подій ймовірність $P_t(k)$ того, що за проміжок часу t подія відбудеться рівно k разів дорівнює

$$P_t(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad (4.9)$$

де λ – інтенсивність потоку.

Приклад 4.7. На телефонну станцію протягом однієї години поступає в середньому 30 викликів. Яка ймовірність того, що протягом хвилини поступить не більше двох викликів?

Розв'язання. Враховуючи, що $1\,\textit{год} = 60\,\textit{хв}$, матимемо $\lambda t = 30 \cdot \frac{1}{60} = \frac{1}{2}$. Шукана ймовірність

$$P(0 \le m \le 2) = P(0) + P(1) + P(2) = \frac{\left(\frac{1}{2}\right)^0}{0!} e^{-\frac{1}{2}} + \frac{\left(\frac{1}{2}\right)^1}{1!} e^{-\frac{1}{2}} + \frac{\left(\frac{1}{2}\right)^2}{2!} e^{-\frac{1}{2}} =$$

$$e^{-\frac{1}{2}}\left(1 + \frac{1}{2} + \frac{1}{8}\right) \approx 0{,}98.$$