Семинар 2

Статистически оценки на параметрите на разпределението

Това са задачи, решаването на които Вас ще задълбочи знания Ви по материала от лекция 2 – плътност на разпределение, функция на разпределение, интегрални граници и пр. Накрая на материала е дадено решението на задачите, но е добре Вие да се опитате да решите задачите самостоятелно.

Задача 2.1*. Дисперсията на една случайна величина X се дефинира като като математическо очакване от квадрата на разликата между случайната величина и нейното математическо очакване.

$$D(X) = M\{[X - M(X)]^2\}$$

Докажете, че горният израз е еквивалентен на

$$D(X) = M(X^2) - M(X)^2. \qquad \text{(II.1)}$$

Задача 2.2*. Докажете, че формулата на квадрата на стандартното отклонение, $\boldsymbol{S^2}$, израз **(2.4)** от лекция 2**,**

$$S^2 = [\Sigma(X_k - \overline{X})^2]/(N-1) \qquad (2.4)$$

може да се преобразува да е равен на:

$$S^2 = (\Sigma X_k{}^2 - N\,\overline{X}^2)/(N-1) \qquad (2.4a)$$

Задача 2.3.** Докажете, че квадратът на стандартното отклонение, $\boldsymbol{S^2}$, е неизместена оценка на дисперсията, докато изразът **(2.5)** от лекция 2 е изместена оценка.

$$S^2 = [\Sigma(X_k - \overline{X})^2]/(N-1) \qquad (2.4)$$

$$(\Sigma X_k^2 - N\,\bar{X}^2)/N \qquad (2.5)$$

Задача 2.4. Програмата Excel има функции, които позволяват намиране на интеграл от плътността на разпределение на различни статистически разпределения, т.е. функцията на разпределение за дадена интегрална граница, както и обратната задача - да се намери интегралната граница при известен интеграл.

Намерете в Help-а описание на следните функции и го прочетете. Попълнете следната таблица. Наименовайте параметрите на функциите с имена от изучавания материал в лекции 1 и 2. Използвайте a) "интегрална граница"; b) "функция на разпределение"; c) "плътност на разпределение"; d) "степени свобода"; e) "едностранна/двустранна постановка"; f) "управляващ параметър".

Функция	**Значение на параметрите**
Y = NORMDIST(x,mean,stand_dev,cum)	Y x mean stand_dev cum
Y = NORMINV(probab,mean,stand_dev)	Y probab mean stand_dev
Y = NORMSINV(probab)	Y probab
Y = CHIDIST(x,deg_freedom)	Y x deg_freedom
Y = CHIINV(probab,deg_freedom)	Y probab deg_freedom
Y = TDIST(x,deg_freedom,tails)	Y x deg_freedom tails

Задача 2.5. Използвайте програмата Excel за да получите решението на задачи 1.10 и 1.11 от лекция 1.

Задача 1.10 (от лекция 1). За стандартното разпределение намерете от приложение 1 **F'(2.87)** и **F(2.87)**, където **F(2.87)** е интеграл в граници от $-\infty$ до 2.87, а **F'(2.87)** - интеграл в граници от -2.87 до +2.87. Използвайте, че **F'(X) = 2F(X) - 1**.

Задача 1.11 (от лекция 1). За нормално разпределение с $\mu = 3.44$ и $\sigma = 0.39$ намерете от приложение 1 **F(4.56)**, където **F(4.56)** е интеграл в граници от $-\infty$ до **4.56**.

Задача 2.6. Използвайте програмата Excel за да получите решението на примери 2.2, 2.3 и 2.4 от лекция 2.

Пример 2.2. (от лекция 2)

a) Да се пресметне вероятността случайната величина χ^2_5 да заема стойности по-малки от 4.3.

b) Да се намери интервалът (0, X), в който случайната величина χ^2_9 заема стойности с вероятност 0.90.

Пример 2.3 (от лекция 2). Да се намери вероятността отношението s^2/σ^2 да е в интервала (0,1), където s е стандартното отклонение от десет измервания, а σ^2 е дисперсията на тяхното разпределение.

Пример 2.4 (от лекция 2). Да се намери вероятността t_9-разпределена случайна величина да заема стойности:

a) по-малки от 2.8

b) в интервала (2.8, 2.8)

Задача 2.7. Подобни функции в програмата Excel има и за разпределението на Фишер. Намерете ги и ги разучете!

Решение на задачите

Задача 2.1*. Дисперсията на една случайна величина X се дефинира като като математическо очакване от квадрата на разликата между случайната величина и нейното математическо очакване.

$$D(X) = M\{[X - M(X)]^2\}$$

Докажете, че горният израз е еквивалентен на

$$D(X) = M(X^2) - M(X)^2. \qquad \text{(II.1)}$$

Решение: Използваме свойствата 1-7 на дисперсията и математическото очакване от лекция 1. Вземаме под внимание, че $M(X)$ е число.

$$D(X) = M\{[X - M(X)]^2\} = M\{X^2 - 2XM(X) + M(X)^2\} =$$

$$= M(X^2) - M\ [2\ M(X)\ X] + M[M(X)^2] =$$

$$= M(X^2) - 2M(X)M(X) + M(X)^2 =$$

$$M(X^2) - M(X)^2$$

Задача 2.2*. Докажете, че формулата на квадрата на стандартното отклонение, ***S***^{2}, израз **(2.4)** от лекция 2,

$$S^2 = [\Sigma(X_k - \bar{X})^2]/(N-1) \qquad (2.4)$$

може да се преобразува да е равен на:

$$S^2 = (\Sigma X_k^2 - N\ \bar{X}^2)/(N-1) \qquad (2.4a)$$

Решение: Преобразуваме числителя на (2.4) по следния начин:

$$\Sigma(X_k - \bar{X})^2 = \Sigma(X_k^2 - 2\ X_k\ \bar{X} + \bar{X}^2) = \Sigma X_k^2 - 2\Sigma X_k\ \bar{X} + \Sigma\ \bar{X}^2 =$$

$$= \Sigma X_k^2 - 2N\ \bar{X}\ \bar{X} + N\ \bar{X}^2 = \Sigma X_k^2 - N\ \bar{X}^2$$

Във времето, когато нямаше компютри, израз (2.4а) се препоръчваше за изчисляване на стандартното отклонение. В него има една сума от квадрати и едно намиране на средната стойност, докато в (2.4) освен тези действия има и изчисляване на N на брой разлики $X_k - \overline{X}$.

Сравнете израз (2.4а) с израз (II.1) от задача 1.

$$D(X) = M(X^2) - M(X)^2. \tag{II.1}$$

$$S^2 = (\Sigma X_k^2 - N\,\overline{X}^2)/(N-1) \tag{2.4a}$$

Задача 2.3.** Докажете, че квадратът на стандартното отклонение, S^2, е неизместена оценка на дисперсията, докато изразът (2.5) от лекция 2 е изместена оценка.

$$S^2 = [\Sigma(X_k - \overline{X})^2]/(N-1) \tag{2.4}$$

$$(\Sigma X_k^2 - N\,\overline{X}^2)/N \tag{2.5}$$

Решение: Една оценка е неизместена, ако математическото и очакване е равно на оценяваната величина, т.е. трябва математическото очакване на случайната величина S^2 да е равно на дисперсията на случайната величина X. Да не забравяме че дисперсията на отделните измервания е еднаква и равна на $D(X) = M\{[(X-M(X)]^2\}$. Трябва да дакажем, че

$$M(S^2) = D(X) = M\{[(X-M(X)]^2\}.$$

В задача 2.2 доказахме, че

$$S^2 = (\Sigma X_k^2 - N\,\overline{X}^2)/(N-1)$$

Тогава

$$M[(N-1)S^2] = M(\Sigma X_k^2 - N\,\overline{X}^2) = M(\Sigma X_k^2) - M(N\,\overline{X}^2) = \Sigma M(X_k^2) - NM(\overline{X}^2) =$$

$$= NM(X^2) - NM(\overline{X}^2)$$

Получихме, че

$$M[(N-1)S^2] = NM(X^2) - NM(\overline{X}^2) \qquad (a)$$

Последният член в (a), $N\,M(\overline{X}^2)$, се развива по следният начин:

$$N\ M(\overline{X}^2) = N\ M\{[(\Sigma X_k)/N]^2\} = N(1/N^2)\ M[(X_1 + X_2 + \ldots + X_N)^2] =$$

$$= (1/N)M(\Sigma X_k^2 + \Sigma X_k X_m) = (1/N)[M(\Sigma X_k^2) + M(\Sigma X_k X_m)] =$$

$$= (1/N)[\Sigma M(X_k^2) + \Sigma\Sigma M(X_k X_m)] \qquad (b)$$

Отделните измервания са независими едно от друго и имат еднакво математическо очакване, затова последният член в скобите на (b) е равен на:

$$\Sigma\Sigma M(X_k X_m) = \Sigma\Sigma M(X_k)\,M(X_m) = \Sigma\Sigma M(X)\,M(X) = \Sigma\Sigma M(X)^2 = N(N-1)M(X)^2$$

Тук използвахме че в смесената сума има $N(N-1)$ членове.

Като отчетем, че първият член в скобите на (b) е равен на

$$\Sigma M(X_k^2) = \Sigma M(X^2) = NM(X^2)$$

за израза (b) получаваме

$$(1/N)[\Sigma M(X_k^2) + \Sigma\Sigma M(X_k X_m)] = (1/N)[NM(X^2) + N(N-1)M(X)^2] =$$

$$= M(X^2) + (N-1)M(X)^2$$

Т.е. изразът (b) е

$$N\ M(\overline{X}^2) = M(X^2) + (N-1)M(X)^2 \qquad (c)$$

След като заместим израза (c) в (a) получаваме (двата израза (b) и (c) са едно и също нещо!)

$$= NM(X^2) - NM(\overline{X}^2) = NM(X^2) - [M(X^2) + (N-1)M(X)^2] =$$

$$= NM(X^2) - M(X^2) - (N-1)M(X)^2 = (N-1)M(X^2) - (N-1)M(X)^2 =$$

$$= (N-1)[M(X^2) - M(X)^2]$$

Т.е. получихме, че

$$M[(N-1)S^2] = (N-1)M[S^2] = (N-1)[M(X^2) - M(X)^2] \qquad (d)$$

Ако разделим израз (d) на $(N-1)$ получаваме

$$M[S^2] = M(X^2) - M(X)^2 \qquad (e)$$

В задача 2.1 получихме израз (II.1)

$$D(X) = M(X^2) - M(X)^2 \qquad (II.1)$$

Приравнявайки изразите (e) и (II.1) получаваме окончателно

$$M[S^2] = D(X) \qquad (f)$$

Ако бяхме използвали формула (2.5), в лявата страна на израза (f) щяхме да получим $N/(N-1)M[S^2]$, от което следва вместо израз (f) ще имаме

$$M[S^2] = [(N-1)/N]D(X)$$

Задача 2.4. Програмата Excel има функции, които позволяват намиране на интеграл от плътността на разпределение на различни статистически разпределения, т.е. функцията на разпределение за дадена интегрална граница, както и обратната задача - да се намери интегралната граница при известен интеграл. Намерете в Help-а описание на следните функции и го прочетете.

Попълнете следната таблица. Наименовайте параметрите на функциите с имена от изучавания материал в лекции 1 и 2. Използвайте a) "интегрална граница"; b) "функция на разпределение"; c) "плътност на разпределение"; d) "степени свобода"; e) "едностранна/двустранна постановка"; f) "управляващ параметър".

‡ Обърнете внимание, че част от функциите работят с нивото на значимост α (интеграл от плътността на разпределение от интегралната граница до безкрайност), а не с функцията на разпределение F(X) (интеграл от плътността на разпределение от минус безкрайност до интегралната граница)! Връзката между тях е очевидно $F(X) = 1 - \alpha$.

Функция	Значение на параметрите
Y = NORMDIST(x,mean,stand_dev,cum)	Y – функция или плътност на разпределение x – интегралната граница mean – математическо очакване stand_dev – стандартно отклонение cum – дали е функция или плътност на разпределение
Y = NORMINV(probab,mean,stand_dev)	Y – интегралната граница probab – нивото на значимост mean – математическо очакване stand_dev – стандартно отклонение
Y = NORMSINV(probab)	Y – интегралната граница probab – нивото на значимост
Y = CHIDIST(x,deg_freedom)	Y – нивото на значимост x – интегралната граница deg_freedom – степени свобода
Y = CHIINV(probab,deg_freedom)	Y – интегралната граница probab – нивото на значимост deg_freedom – степени свобода
Y = TDIST(x,deg_freedom,tails)	Y – нивото на значимост x – интегралната граница deg_freedom – степени свобода tails – постановка на задачата: двустранна или едностранна

Отворете файла **seminar02_distributions.xls** и в таблицата "**problem 2.4**" сравнете прилагането на **Excel** функцията **CHIDIST()** с различни интегрални граници. Вижте, че има разлика от приложението на функцията **NORMDIST()**, която е изследвана в таблицата "**NORMDIST**". Следващите три задачи също са решени в този файл, но се опитайте да ги решите сами, като използвате таблицата по-горе.

Задача 2.5. Използвайте програмата Excel за да получите решението на задачи 1.10 и 1.11 от лекция 1. Отворете файла **seminar02_distributions.xls** и разгледайте решенията в него.

Задача 1.10 (от лекция 1). За стандартното разпределение намерете от приложение 1 **F'(2.87)** и **F(2.87)**, където **F(2.87)** е интеграл в граници от $-\infty$ до 2.87, а F'(2.87) - интеграл в граници от -2.87 до +2.87. Използвайте, че F'(X) = 2F(X) - 1.

Решение: За F(2.87) функцията е NORMDIST(2.87,0,1,TRUE), която дава същия резултат, 0.99795, като този, получен от приложение 1. Същият резултат се получава и с функцията на стандартното разпределение, NORMSDIST(2.87). Няма функция за двустранно постановка на въпроса, затова **F'(2.87)** се намира като в решението на задача 1.10:

F'(2.87) = 2xF(2.87) - 1 = 2x0.99795 - 1 = 0.9959

Вместо константата TRUE може да се запише единица, а вместо FALSE – нула: в този случай функцията връща плътността на разпределение - вижте таблицата NORMDIST в същия файл.

Задача 1.11 (от лекция 1). За нормално разпределение с μ = 3.44 и σ = 0.39 намерете от приложение 1 F(4.56), където F(4.56) е интеграл в граници от $-\infty$ до 4.56.

Решение: За F(4.56) функцията е NORMDIST(4.56,3.44,0.39,TRUE), която дава същия резултат, 0.99796, като този, получен от приложение 1. Обърнете внимание, че числата се различават в петия знак, след десетичната запетая - програмата Excel изчислява интеграла на Лаплас с приближени числени методи - спомнете си, че той няма аналитично решение за произволни граници!

Задача 2.6. Използвайте програмата Excel за да получите решението на примери 2.2, 2.3 и 2.4 от лекция 2.

Пример 2.2 (от лекция 2).

а) Да се пресметне вероятността случайната величина χ^2_5 да заема стойности по-малки от 4.3.

Решение: Нивото на значимост се изчислява с `CHIDIST(4.5,5)` – получава се **0.48**. За функцията на разпределения (това е вероятността, която се търси) се получава, че е 1 - **0.48** = **0.52**, която е близка до 0.50, намерено от приложение 2 . Обърнете внимание, че в таблицата имаме само краен брой числа за интегралната граница (числото 4.5 го няма), докато програмата Excel изчислява функцията на разпределение с произволни интегрални граници: изчисленията не само са по-бързи от търсене в таблица, но и по-точни.

b) Да се намери интервалът `(0, X)`, в който случайната величина χ^2_9 заема стойности с вероятност 0.90.

Решение: Функцията е `CHIINV(0.1,9)`, която дава близък резултат, 14.68, до 14.7, намерено от приложение 2. Обърнете внимание, че във функцията под името `probability` стои нивото на значимост `0.1`, а не функцията на разпределение `0.9`.

Пример 2.3 (от лекция 2). Да се намери вероятността отношението s^2/σ^2 да е в интервала (0,1), където s е стандартното отклонение от десет измервания, а σ^2 е дисперсията на тяхното разпределение.

Решение: Нивото на значимост се изчислява с `CHIDIST(9,9)` и е **0.44**. Функцията е 1 - **0.44** = **0.56**, която е близка до намерената 0.50 от приложение 2. Обърнете внимание, че разминаването в стойностите е повече от 10%, затова във времената, когато нямаше Excel при работа с интегрална граница, която я няма в таблицата можеше да се направи линейна интерполация.

Пример 2.4 (от лекция 2). Да се намери вероятността t_9-разпределена случайна величина да заема стойности:

а) по-малки от 2.8

Решение: Нивото на значимост се изчислява с `TDIST(2.8,9,1)` и е **0.01036**. Функцията е 1 - **0.01036** = **0.98964**, която е близка до намерената 0.99 от приложение 3. Обърнете внимание, че отново Excel функцията дава нивото на значимост, а не функцията на разпределение.

b) в интервала (2.8, 2.8)

Решение: Нивото на значимост се изчислява с `TDIST(2.8,9,2)` и е **0.02071**. Функцията е 1 - **0.02071** = **0.97929**, която е близка до намерената 0.98 от приложение 3. Разликата с предишната точка е, че `tails = 2`, а не `1`. Както се очакваше, поради симетричността на разпределението нивото на значимост е двойно по-голямо от това в т. а). Обърнете внимание, че отново Excel функцията дава нивото на значимост, а не функцията на разпределение.