Задание

1. Изучить и кратко законспектировать материал лекции. Обязательно решить задания в лекции.
2. Фотоотчет присылать на электронную почту

С уважением, Хвастова Светлана Ивановна

!!! Если возникнут вопросы обращаться по телефону 0721389311.

Электронная почта: xvsviv@rambler.ru

Распределение Пуассона

Цель: изучить параметры распределения Пуассона, выучить формулы определения вероятности случайной величины распределения Пуассона и его числовых характеристик.

Ранее мы рассмотрели два типа дискретных числовых распределений: биномиальное и гипергеометрическое. Во многих практически важных приложениях большую роль играет распределение Пуассона. Многие из числовых дискретных величин являются реализациями пуассоновского процесса, обладающего следующими свойствами:[1]

- Нас интересует, сколько раз происходит некое событие в заданной области возможных исходов случайного эксперимента. Область возможных исходов может представлять собой интервал времени, отрезок, поверхность и т.п.
- Вероятность данного события одинакова для всех областей возможных исходов.
- Количество событий, происходящих в одной области возможных исходов, не зависит от количества событий, происходящих в других областях.
- Вероятность того, что в одной и той же области возможных исходов данное событие происходит больше одного раза, стремится к нулю по мере уменьшения области возможных исходов.

Чтобы глубже понять смысл пуассоновского процесса, предположим, что мы исследуем количество клиентов, посещающих отделение банка,

[1] Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 320–328

расположенное в центральном деловом районе, во время ланча, т.е. с 12 до 13 часов. Предположим, требуется определить количество клиентов, приходящих за одну минуту. Обладает ли эта ситуация особенностями, перечисленными выше? Во-первых, событие, которое нас интересует, представляет собой приход клиента, а область возможных исходов — одноминутный интервал. Сколько клиентов придет в банк за минуту — ни одного, один, два или больше? Во-вторых, разумно предположить, что вероятность прихода клиента на протяжении минуты одинакова для всех одноминутных интервалов. В-третьих, приход одного клиента в течение любого одноминутного интервала не зависит от прихода любого другого клиента в течение любого другого одноминутного интервала. И, наконец, вероятность того, что в банк придет больше одного клиента стремится к нулю, если временной интервал стремится к нулю, например, становится меньше 0,1 с. Итак, количество клиентов, приходящих в банк во время ланча в течение одной минуты, описывается распределением Пуассона.

Распределение Пуассона имеет один параметр, обозначаемый символом λ (греческая буква «лямбда») – среднее количество успешных испытаний в заданной области возможных исходов. Дисперсия распределения Пуассона также равна λ, а его стандартное отклонение равно $\sqrt{\lambda}$. Количество успешных испытаний X пуассоновской случайной величины изменяется от 0 до бесконечности. Распределение Пуассона описывается формулой:

(1) $P(X) = \frac{e^{-\lambda}\lambda^{X}}{X!}$

где $P(X)$ — вероятность X успешных испытаний, λ — ожидаемое количество успехов, e— основание натурального логарифма, равное 2,71828, X— количество успехов в единицу времени.

Вернемся к нашему примеру. Допустим, что в течение обеденного перерыва в среднем в банк приходят три клиента в минуту. Какова вероятность того, что в данную минуту в банк придут два клиента? А чему равна вероятность того, что в банк придут более двух клиентов?

Применим формулу (1) с параметром $X = 3$. Тогда вероятность того, что в течение данной минуты в банк придут два клиента, равна

$$P(X=2)=\frac{e^{-3,0}(3,0)^2}{2!}=\frac{9}{(2,71828)^3\times 2}=0,2240$$

Вероятность того, что в банк придут более двух клиентов, равна P(X > 2) = P(X = 3) + P(X = 4) + ... + P(X = ∞) . Поскольку сумма всех вероятностей должна быть равной 1, члены ряда, стоящего в правой части формулы, представляют собой вероятность дополнения к событию X≤ 2. Иначе говоря, сумма этого ряда равна 1 – P(X ≤ 2). Таким образом, P(X> 2) = 1 – P(X≤2) = 1 – [P(X = 0) + P(X = 1) + P(X = 2)]. Теперь, используя формулу (1), получаем:

$$P(X>2)=1-\left[\frac{e^{-3,0}(3,0)^0}{0!}\right]+\left[\frac{e^{-3,0}(3,0)^1}{1!}\right]+\left[\frac{e^{-3,0}(3,0)^2}{2!}\right]=0,5768$$

Таким образом, вероятность того, что в банк в течение минуты придут не больше двух клиентов, равна 0,423 (или 42,3%), а вероятность того, что в банк в течение минуты придут больше двух клиентов, равна 0,577 (или 57,7%).

Такие вычисления могут показаться утомительными, особенно если параметр λ достаточно велик. Чтобы избежать сложных вычислений, многие пуассоновские вероятности можно найти в специальных таблицах (рис. 1). Например, вероятность того, что в заданную минуту в банк придут два клиента, если в среднем в банк приходят три клиента в минуту, находится на пересечении строки X = 2 и столбца λ = 3. Таким образом, она равна 0,2240 или 22,4%.

	λ			
X	2.1	2.2	...	3.0
0	0,1125	0,1108	...	0,0498
1	0,2572	0,2438	...	0,1494
2	0,2700	0,2681	...	0,2240
3	0,1890	0,1966	...	0,2240
4	0,0992	0,1082	...	0,1680
5	0,0417	0,0476	...	0,1008
6	0,0146	0,0174	...	0,0504
7	0,0044	0,0055	...	0,0216
8	0,0011	0,0015	...	0,0081
9	0,0003	0,0004	...	0,0027
10	0,0001	0,0001	...	0,0008
11	0,0000	0,0000	...	0,0002
12	0,0000	0,0000	...	0,0001

Рис. 1. Пуассоновская вероятность при λ = 3

Сейчас вряд ли кто-то будет пользоваться таблицами, если под рукой есть Excel с его функцией =ПУАССОН.РАСП() (рис. 2). Эта функция имеет три параметра: число успешных испытаний *X*, среднее ожидаемое количество успешных испытаний λ, параметр *Интегральная*, принимающий два значения: ЛОЖЬ – в этом случае вычисляется вероятность числа успешных испытаний *X* (только X), ИСТИНА – в этом случае вычисляется вероятность числа успешных испытаний от 0 до *X*.

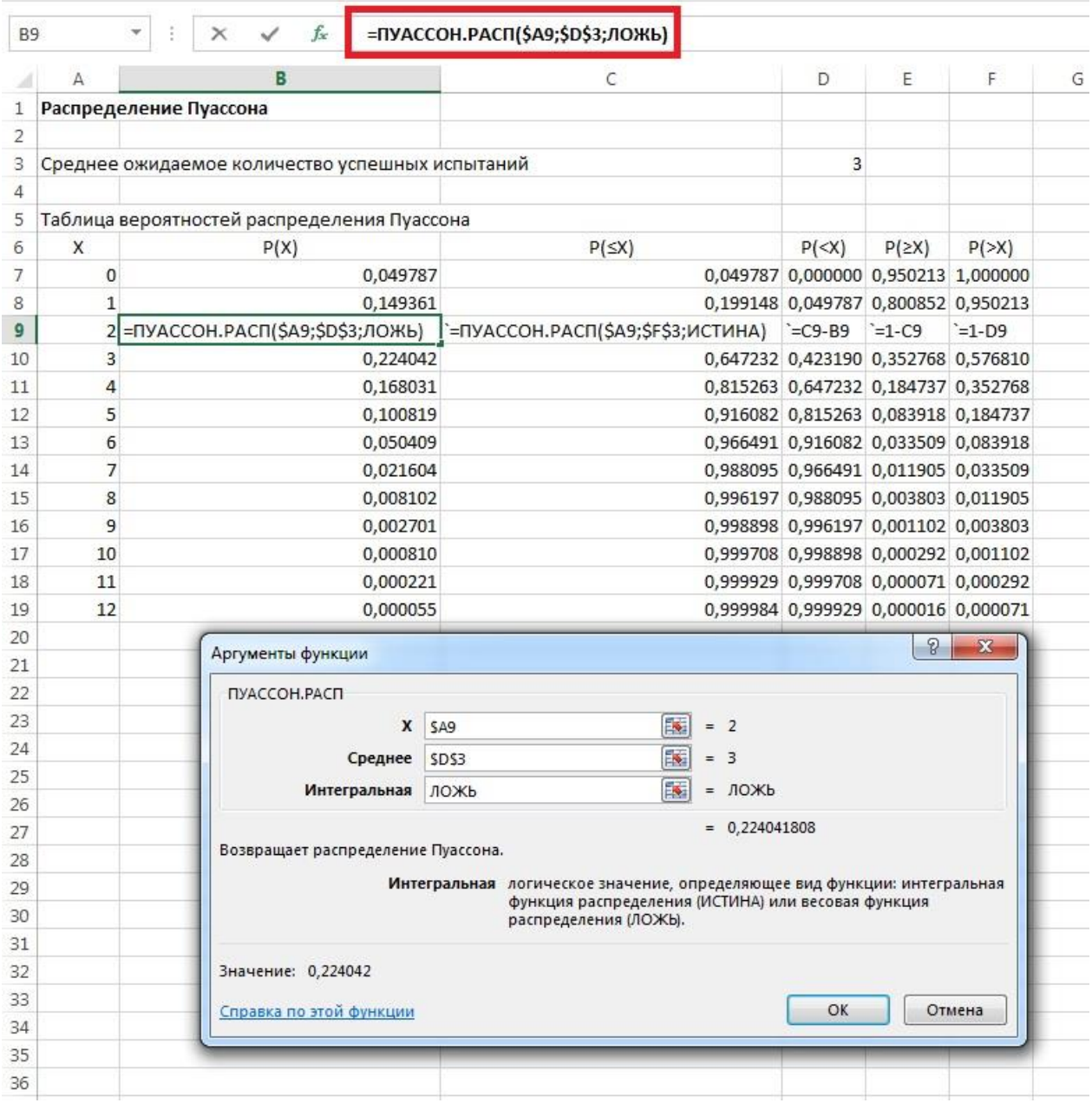

Рис. 2. Расчет в Excel вероятностей распределения Пуассона при λ = 3

Аппроксимация биноминального распределения с помощью распределения Пуассона

Если число n велико, а число p — мало, биномиальное распределение можно аппроксимировать с помощью распределения Пуассона. Чем больше число n и меньше число p, тем выше точность аппроксимации. Для аппроксимации биномиального распределения используется следующая модель Пуассона.

(2) $P(X) \cong \frac{e^{-np}(np)^X}{X!}$

где $P(X)$ — вероятность X успехов при заданных параметрах n и p, n — объем выборки, p— истинная вероятность успеха, e— основание натурального логарифма, X — количество успехов в выборке (X = 0, 1, 2, ..., n).

Теоретически случайная величина, имеющая распределение Пуассона, принимает значения от 0 до ∞. Однако в тех ситуациях, когда распределение Пуассона применяется для приближения биномиального распределения, пуассоновская случайная величина — количество успехов среди n наблюдений — не может превышать число n. Из формулы (2) следует, что с увеличением числа n и уменьшением числа p вероятность обнаружить большое количество успехов уменьшается и стремится к нулю.

Как говорилось выше, математическое ожидание μ и дисперсия σ^2 распределения Пуассона равны λ. Следовательно, при аппроксимации биномиального распределения с помощью распределения Пуассона для приближения математического ожидания следует применять формулу (3).

(3) $\mu = E(X) = \lambda = np$

Для аппроксимации стандартного отклонения используется формула (4).

(4) $\sigma = \sqrt{\lambda} = \sqrt{np}$

Обратите внимание на то, что стандартное отклонение, вычисленное по формуле (4), стремится к стандартному отклонению в биномиальной модели – $\sigma = \sqrt{np(1-p)}$, когда вероятность успеха p стремится к нулю, и, соответственно, вероятность неудачи $1 - p$ стремится к единице.

Предположим, что 8% шин, произведенных на некотором заводе, являются бракованными. Чтобы проиллюстрировать применение распределения Пуассона для аппроксимации биномиального распределения, вычислим вероятность

обнаружить одну дефектную шину в выборке, состоящей из 20 шин. Применим формулу (2), получим

$$P(X=1) \cong \frac{e^{-20*0,08}(20*0,08)^1}{1!} = 0,3230$$

Если бы мы вычислили истинное биномиальное распределение, а не его приближение, то получили бы следующий результат:

$$p(X=1) = \binom{20}{1} \times 0,08 \times (0,92)^{19} = 0,3282$$

Однако эти вычисления довольно утомительны. В то же время, если вы используете Excel для вычисления вероятностей, то применение аппроксимации в виде распределения Пуассона становится излишним. На рис. 3 показано, что трудоемкость вычислений в Excel одинакова. Тем не менее, этот раздел, на мой взгляд, полезен понимаем того, что при некоторых условиях биноминальное распределение и распределение Пуассона дают близкие результаты.

	A	B	C	D
1	**Расчет шин**			
2				
3	*На основе аппроксимации биноминального распределения распределением Пуассона*			
4	Число шин	n	20	
5	Верочтность брака	p	8%	
6	Среднее ожидаемое количество успешных испытаний	λ	1,6	
7	Вероятность встретить бракованную шину в выборе из	X	1	
8			Расчет	Использованная формула
9	Распределение Пуассона		0,3230	`=ПУАССОН.РАСП(C7;C6;ЛОЖЬ)
10				
11	*На основе биноминального распределения (точно, а не приближенно)*			
12	Число шин	n	20	
13	Верочтность брака	p	8%	
14	Вероятность встретить бракованную шину в выборе из	X	1	
15			Расчет	Использованная формула
16	Биноминальное распределение		0,3282	`=БИНОМ.РАСП(C14;C12;C13;ЛОЖЬ)
17				

Рис. 3. Сравнение трудоемкости расчетов в Excel: (а) распределение Пуассона; (б) биноминальное распределение

Итак, в настоящей и двух предыдущих заметках были рассмотрены три дискретных числовых распределения: биномиальное, гипергеометрическое и Пуассона. Чтобы лучше представлять, как эти распределения соотносятся друг с другом приведем небольшое дерево вопросов (рис. 4).

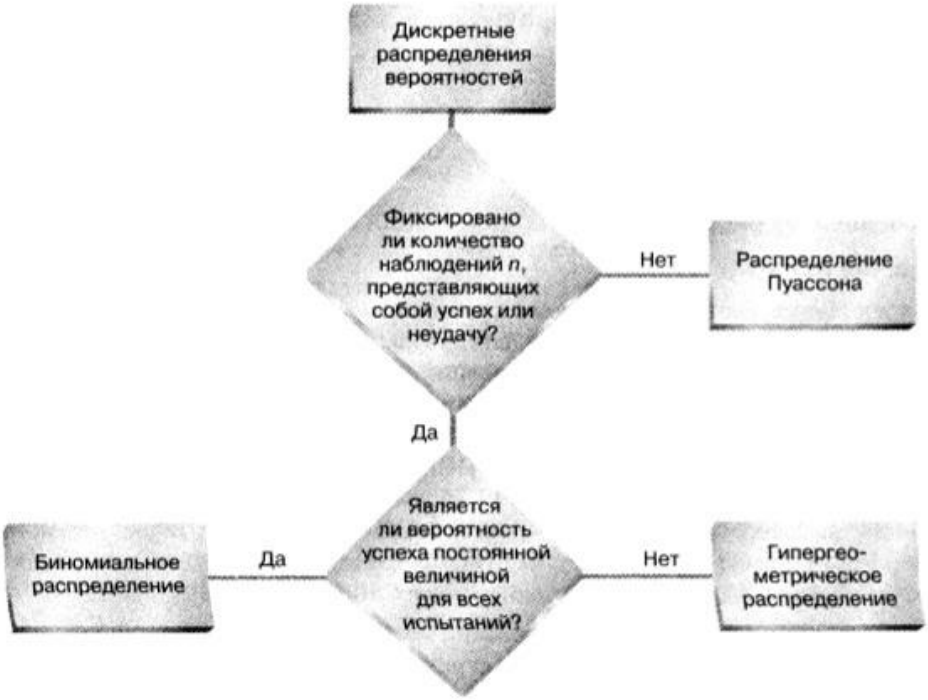

Рис. 4. Классификация дискретных распределений вероятностей

Предыдущая заметка Гипергеометрическое распределение

Следующая заметка

К оглавлению Статистика для менеджеров с использованием Microsoft Excel