Теория вероятностей

Случайные величины

Черновик

Кафедра СМиМ

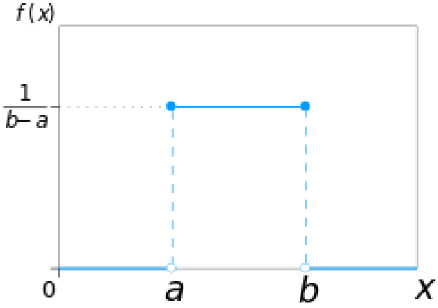
2019

План

Outline

Outline

Равномерное распределение



Все возможные значения случайной величины равновероятны.

Равномерное распределение может иметь как дискретная случайная величина, так и непрерывная

`scipy.stats.uniform.rvs ( loc = a, scale = b)`

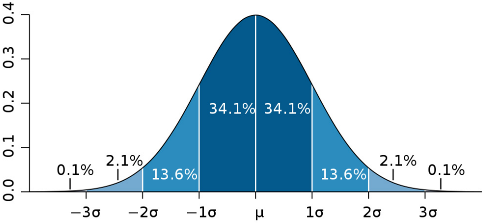
Равномерное распределение. Примеры

- Количество очков, выпавших на игральной кости
- Число выпавшее на рулетке
- Номер автобусного билета (в единичном испытании)
- Время ожидания события, происходящего со строгой периодичностью. например время ожидания поезда, который отправляется со станции раз в 30 минут

Значения случайной величины с равномерным распределением используются для осуществления случайных выборок.

Outline

Нормальное распределение



μ - математическое ожидание, σ - среднеквадратичное отклонение.

Возможные значения СВ близкие к мат. ожидания наиболее вероятны.

Если СВ является суммой большого числа других независимых величин, то она подчинятся нормальному закону распределения.[1]

[1] см. центральная предельная теорема

Нормальное распределение. Примеры

- Рост человека
- Ошибка измерения
- Прочность бетона
- Масса новорождённых детей
- Объём молока производимый коровой каждый день

Нормальное распределение

- Функция распределения

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t$$

- Параметры
  - μ - математическое ожидание
  - σ - стандартное отклонение

Стандартное нормальное распределение

- При $\mu = 0$ и $\sigma = 1$ распределение называется *стандартным нормальным распределением*
- Нормирование случайной величины:

$$z = \frac{x - \mu}{\sigma}$$

  где x - исходное значение случайной величины;
  z - нормированное значение.
- Тогда функция распределения

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} \, \mathrm{d}t$$

Нормальное распределение и функция Лапласа

- В таблицах может приводится значение функции, где нижний предел 0 вместо $-\infty$

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} \,\mathrm{d}t$$

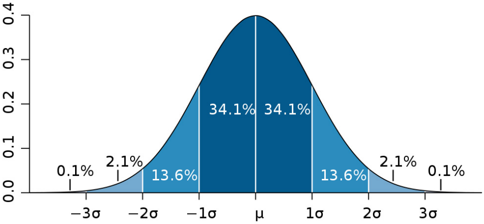
- Чтобы перейти от $F_0(x)$ к $F(X)$:

$$F(X) = 0.5 + F_0(X)$$

- 0.5 соответствует площади под кривой слева от нуля

Правило трёх сигм

Вероятность того, что случайная величина отклонится от своего математического ожидания на большую величину, чем утроенное среднее квадратичное отклонение, практически равна нулю.



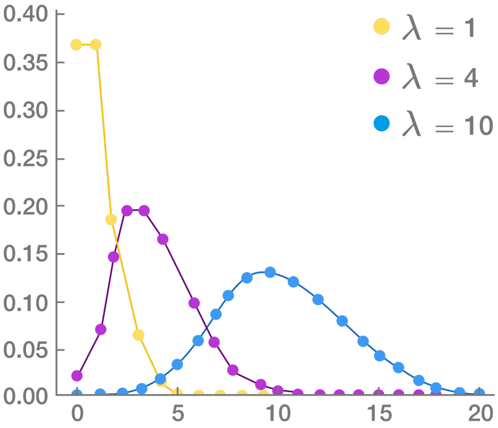
$$P(\mu - 1\sigma \leq X \leq \mu + 1\sigma) \approx 0.6827$$

$$Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545$$

$$Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973$$

Outline

Распределение Пуассона



СВ - количество событий на меру пространства или времени, при средней частоте λ

Распределение Пуассона. Примеры

ДТП в определённом районе города случается в среднем дважды в неделю.

Какова вероятность того, что на этой неделе не будет ДТП?

Используем закон Пуассона[2]:

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

где $\lambda = 2$, $k = 0$ - число событий.

тогда $P(0) = \frac{2^0 e^{-2}}{0!} = 0.14$

[2] в пакете scipy: `scipy.stats.distributions.poisson.pmf(x, lambda)`

Распределение Пуассона. Примеры

ДТП в определённом районе города случается в среднем дважды в неделю.

Какова вероятность того, что на этой неделе не будет ДТП?

Используем закон Пуассона[2]:

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

где $\lambda = 2$, $k = 0$ - число событий.

тогда $P(0) = \frac{2^{\mathbf{0}} e^{-\mathbf{2}}}{0!} = 0.14$

Какова вероятность того, что на этой неделе будет 1 и 2 ДТП?
$P(1) = \frac{2^{\mathbf{1}} e^{-\mathbf{2}}}{1!} = 0.27$, $P(2) = \frac{2^{\mathbf{2}} e^{-\mathbf{2}}}{2!} = 0.27$

Какова вероятность того, что на этой неделе будет больше 2-х ДТП?

$P (X > 2) = 1 - P(0) - P(1) - P(2)$

[2] в пакете scipy: `scipy.stats.distributions.poisson.pmf(x, lambda)`

Распределение Пуассона. Примеры

Для при определении вероятности для заданного числа событий произошедших за t единиц времени параметр λ определяют так:

$$\lambda = tn$$

где n число событий за единицу времени

Сравним вероятности следующих событий:

- 3 ДТП за неделю
- 15 ДТП за 5 недель

Распределение Пуассона. Примеры

Для при определении вероятности для заданного числа событий произошедших за t единиц времени параметр λ определяют так:

$$\lambda = tn$$

где n число событий за единицу времени

Сравним вероятности следующих событий:

- 3 ДТП за неделю
- 15 ДТП за 5 недель

$n = 2,\ t_1 = 1,\ \lambda_1 = 2,\ t_2 = 5,\ \lambda_2 = 2 \cdot 5 = 10$

$P(3, \lambda_1 = 2) = 0.18$
$P(15, \lambda_2 = 10) = 0.035$

Распределение Пуассона. Примеры

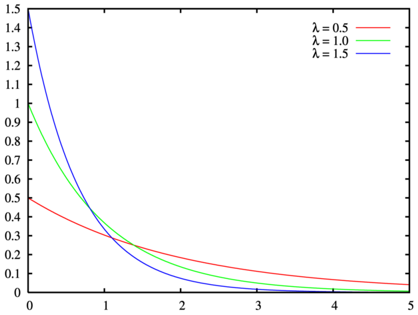
Величины подчиняющиеся распределению Пуассона

- Число изюминок в булочке
- Число мутация в ДНК
- Число звонков в службу технической поддержки
- Число смертей в год для заданной возрастной категории
- Число альфа-частиц излучённых за определённый промежуток времени

kvant.mccme.ru/1988/08/raspredelenie_puassona.htm

Outline

Экспоненциальное (показательное) распределение



λ - скорости. $M(X) = \frac{1}{\lambda}$

моделирует время между двумя последовательными свершениями одного и того же события.

Экспоненциальное (показательное) распределение. Пример

Среднее время ожидания покупателя - 15 минут. Какова вероятность, что во время перерыва длительностью 5, 10 и 15 минут придёт покупатель?

Тогда $\frac{1}{\lambda} = 15 \to \lambda = 0.067$.

$$P(X < 5) = F(5) = 1 - e^{-0.067 \cdot 5} = 0.28$$
$$P(X < 10) = F(10) = 1 - e^{-0.067 \cdot 10} = 0.49$$
$$P(X < 15) = F(15) = 1 - e^{-0.067 \cdot 15} = 0.63$$

```
scipy.stats.expon.cdf ( x = 5, scale = 15)  # 0.28
```

Экспоненциальное распределение vs распределение Пуассона

В чём разница и что общее у экспоненциального распределения и распределения Пуассона?

Экспоненциальное (показательное) распределение

Величины подчиняющиеся экспоненциальному распределению

- Расстояние между участками ДНК с мутациями
- Время ожидания звонка службу технической поддержки
- Время между излучением частиц
- Расстояние между местами ДТП на дороге[3]

[3] в предположении, что вероятность ДТП на каждом участке дороги одинакова

Outline

Распределение Вейбулла

.

Outline

Распределение Гумбеля

.

Outline

Нелинейность - это сложно

- Система из независимых случайных величин $X_1, X_2, ..., X_n$
- Для каждой из величин известны математическое ожидание m_{x1} и стандартное отклонение σ_{x1}
- Рассмотрим новую случайную величину Y, которая зависит от $X_1, X_2, ..., X_n$

$$Y = f(X_1, X_2, ..., X_n)$$

- Требуется определить математическое ожидание и стандартное отклонение с.в. Y

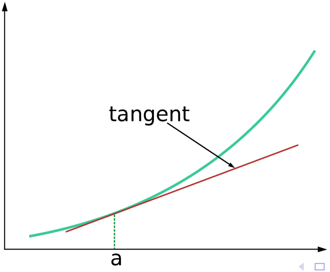
Функция случайного аргумента

- Определение математического ожидания с.в. требует знания функции плотности, а так же интегрирования этой функции.
- Эта процедура может быть сложной, кроме того некоторые интегралы невозможно вычислить аналитически
- Однако если функция f линейна то математическое ожидание и дисперсия вычисляются просто
- Даже если f нелинейна её можно рассматривать как линейную на небольшой области определения
- Далее для простоты рассмотрим функцию для f одного аргумента

$$Y = f(X)$$

Линейная аппроксимация

- Нелинейную функцию можно аппроксимировать линейной в окрестности некоторой точки
- Аппроксимирующую прямую можно выбрать совпадающую с касательной к функции в этой точке
- Если предполагается, что рассматриваемая окрестность точки, через которую проведена касательная небольшая, то и ошибка аппроксимации может быть невелика



Ряд Тейлора

- Для аппроксимации используем разложении функции в ряд Тейлора в окрестности точки a

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(i)}(a)}{n!}(x-a)^n$$

$f^{(n)}(a)$ - значение n-й производной функции f в точке a

- Запишем только первые два слагаемых ряда

$$f(x) = \frac{f(a)}{0!}(x-a)^0 + \frac{f'(a)}{1!}(x-a)^1$$

Ряд Тейлора

- Для аппроксимации используем разложении функции в ряд Тейлора в окрестности точки a

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(i)}(a)}{n!}(x-a)^n$$

$f^{(n)}(a)$ - значение n-й производной функции f в точке a

- Запишем только первые два слагаемых ряда

$$f(x) = \frac{f(a)}{0!}(x-a)^0 + \frac{f^{'}(a)}{1!}(x-a)^1$$

$$f(x) \approx f(a) + f^{'}(a) \cdot (x-a)$$

- Это уравнение касательной к функции f в точке a

Ряд Тейлора

Пример

Разложим функцию $\sin(x)$ в окрестности 0

Ряд Тейлора

Пример

Разложим функцию sin(x) в окрестности 0

$$sin(x) \approx 0 + 1 \cdot (x - 0) = x$$

x	*sin x*	*tailor*	*delta*
0,000	0,000	0,000	0,000
0,001	0,001	0,001	0,000
0,010	0,010	0,010	0,000
0,100	0,100	0,100	0,000
0,200	0,199	0,200	0,001
0,300	0,296	0,300	0,004
0,400	0,389	0,400	0,011

Значения x приведено в радианах

Разложим функцию $y = f(x)$ в окрестности математического ожидания m_x:

$$y = f(m_x) + f'(m_x) \cdot (x - m_x)$$

Найдём найдём математическое ожидание y:

$$m_y = M[f(m_x) + f'(m_x) \cdot (x - m_x)] = M[f(m_x)] = f(m_x)$$

Найдём найдём дисперсию y используя формулу со слайда **??**:

$D_y = M[[f(m_x) + f'(m_x) \cdot (x - m_x) - (f(m_x) + f'(m_x) \cdot (x - m_x)]^2] = M[[f'(m_x)]^2 \cdot (x - m_x)^2] = [f'(m_x)]^2 \sigma_x^2$

Таким образом, чтобы определить математическое ожидание и дисперсию линеаризованной функции случайного аргумента достаточно знать:

- Математическое ожидание аргумента
- Дисперсию аргумента

$$m_y = f(m_x)$$

$$D_y = [f'(m_x)]^2 \sigma_x^2$$

Математическое ожидание и дисперсия функции нескольких переменных

Если функция зависит от нескольких переменных $x_1, x_2, ..., x_n$

$$m_y = f(m_{x1}, m_{x2}, ..., m_{xn})$$

$$D_y = \sum(\frac{\partial f(m_{x1}, m_{x2}, ..., m_{xn})}{\partial x_i})^2 \sigma_{xi}^2 \tag{1}$$

Пример

Относ бомбы выражается приближенной аналитической формулой:

$$X = v_0\sqrt{\frac{2H}{g}}(1 - 1.8 \cdot 10^{-5})cH$$

где v_0 - скорость самолета (м/с), H - высота сбрасывания (м), c - баллистический коэффициент.

По приборам определены: H $= 4000$, $\sigma_H = 40$ м; $v_0 = 150$ м/с, $\sigma_{v_0} = 1$ м/с; c $= 1$, $\sigma = 0.05$. Ошибки приборов независимы друг от друга.

- Найти относ и среднее квадратическое отклонение точки падения бомбы вследствие неточности в определении параметров v_0, H и c.
- Определить, какой из этих факторов оказывает наибольшее влияние на разброс точки падения бомбы.

Пример

Можно убедится, что при небольших изменениях параметров v_0, H и c функция определяющая относ бомбы остаётся практически линейной

Поэтому замена формул для математического ожидания и стандартного отклонения на аналогичные для линейной аппроксимации оправдана

Пример

- Заданные значения величин v_0, H и c являются средними значениями, так как их отклонения в обе стороны равновероятны

Определим среднее значения относа X:

$$X = 150 \cdot \sqrt{\frac{8000}{9.81} \cdot (1 - 1.8^{-5}) \cdot 1 \cdot 4000} = 3975.12\text{м}$$

Пример

- Определим величину, ошибка определения которой вносит наибольший вклад в величину относа бомбы X
- Для этого вычислим все слагаемые, из которых образуется дисперсия искомой величины (формула 1)
- Сначала определим производные

$$\frac{\partial X}{\partial H} = \frac{v_0}{\sqrt{2Hg}}(1 - 1.8 \cdot 10^{-5})cH - v_0\sqrt{\frac{2H}{g}}(-1.8 \cdot 10^{-5})c$$

Пример

- Определим величину, ошибка определения которой вносит наибольший вклад в величину относа бомбы X
- Для этого вычислим все слагаемые, из которых образуется дисперсия искомой величины (формула 1)
- Сначала определим производные

$$\frac{\partial X}{\partial H} = \frac{v_0}{\sqrt{2Hg}}(1 - 1.8 \cdot 10^{-5})cH - v_0\sqrt{\frac{2H}{g}}(-1.8 \cdot 10^{-5})c$$

$$\frac{\partial X}{\partial v_0}\cdot = \sqrt{\frac{2H}{g}}(1 - 1.8 \cdot 10^{-5})cH$$

$$\frac{\partial X}{\partial c}\cdot = v_0\sqrt{\frac{2H}{g}}(1 - 1.8 \cdot 10^{-5})H$$

Пример

$$(\frac{\partial X}{\partial H} \cdot \sigma_h)^2 = 0.429^2 \cdot 40^2$$

$$(\frac{\partial X}{\partial v_0} \cdot \sigma_{v_0})^2 = 26.4^2 \cdot 1^2$$

$$(\frac{\partial X}{\partial c} \cdot \sigma_c)^2 = (-307)^2 \cdot 0.05^2$$

Пример

Если предположить, что X имеет нормальное распределение, то с какой вероятностью бомба упадёт на далее чем в x м от расчетного места падения?

Outline

Источники

- Теория вероятностей и математическая статистика. Гмурман В.Е. biblio-online.ru/book/teoriya-veroyatnostey-i-matematicheskaya-statistika-431095
- Руководство к решению задач по теории вероятностей и математической статистике. В. Е. Гмурман. — 11-е изд., Издательство Юрайт, 2019. — 406 с www.biblio-online.ru/book/02E0C1D3-4EEA-43AA-AA6B-5E25C4991D0

Ссылки

Материалы курса

github.com/VetrovSV/ST