Анализ социальных сетей. Лекция 3
Сетевые математические модели

Михаил Пожидаев

14 ноября 2024 г.

Идея моделирования

Что делать, если сеть доступна частично ?

Если наблюдается только часть элементов сети (узлы и связи между ними), то необходимо построить модель, максимально точно соответствующей наблюдаемым параметрам и оценки вычислять уже на основе модели, заменяя ею недостающую часть сети.

Роль моделирования:

- обоснование наблюдаемых явлений и оценка влияния различных параметров;
- предсказаний явлений в будущем.

Пример

Наблюдая историю распространения инфекции, можно прогнозировать развитие эпидемии в будущем и планировать карантинные мероприятия.

Нормальное распределение

И центральная предельная теорема

$$p_{\mu\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Параметры:

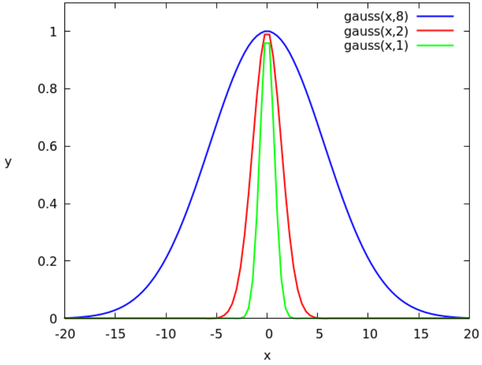
- μ — математическое ожидание (сдвиг вправо-влево);
- σ — дисперсия (толщина «колокола»).

Центральная предельная теорема

Сумма распределений вероятностей большого количества равных независимых случайных величин стремится к нормальному распределению.

Нормальные распределения

С разной дисперсией



Примеры

Какие явления подчиняются нормальному закону?

1. Случайные ошибки измерений.
2. Угол наклона морского пляжа.
3. Размеры беспозвоночных ископаемых организмов.
4. Уровень воды в скважине в течение времени.
5. Результаты испытаний стали на прочность.
6. Топография, плотность морского песка.
7. Показатель окатанности галек разного размера.
8. Плотность заполнения зерен в песчанике.
9. Угол склона речной долины.
10. Содержание влаги в осадочных породах.

Степенное распределение

Распределение Парето или распределение Брэдфорда

$$p(x) = cx^{-\alpha},$$

где α — ключевой параметр распределения, требующий правильного выбора.

Свойства:

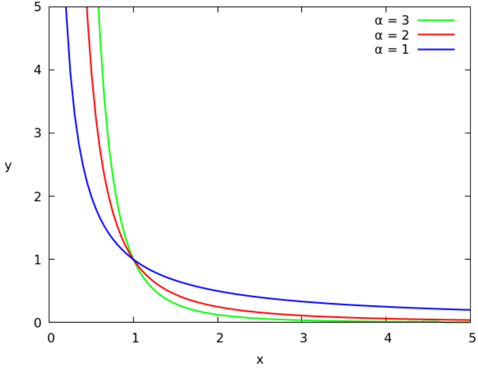
- для малых x не имеет смысла: $x >= x_{min} > 0$;
- любой закон распределения требует нормировки — общая вероятность всегда равна единице: $\int\limits_{x_{min}}^{\infty} cx^{-\alpha} dx = 1$ (в дискретном случае интеграл заменяется суммой).

В результате:

выполнение обоих свойств позволяет определить константу $c = (\alpha - 1){x_{min}}^{\alpha-1}$ в непрерывном случае и $c = \frac{1}{\sum\limits_{i > x_{min}} i^{-\alpha}}$ в дискретном случае.

$y = x^{-\alpha}$

Степенное распределение



Свойства α

Зависимость распределения от показателя степени

При $\alpha \leq 1$:

распределения не существует, потому что не выполняется условие нормировки: $\int \frac{1}{\alpha} dx = \ln|x|$, а $\lim\limits_{x \to \infty} \ln|x| = \infty$.

При $\alpha \leq 2$:

распределение существует, но его математическое ожидание стремится к бесконечности (при любых наблюдениях случайной величины в прошлом новое наблюдение может оказаться значительно больше).

При $\alpha \leq 3$:

распределение существует, и математическое ожидание конечно, но дисперсия бесконечна.

Во всех остальных случаях существуют конечные математическое ожидание и дисперсия.

Как найти α?

Используем метод максимального правдоподобия

При известном множестве из n наблюдений x_i оценка влияния параметра α выражается следующей функцией:

$$L(\alpha) = \sum_{i=1}^{n} -\alpha \ln x_i + n \ln(\alpha - 1) + n(\alpha - 1) \ln x_{min}$$

Дифференцируя это выражение ($\frac{dl}{d\alpha}$) и приравняв её значение к нулю, получим выражение для искомой $\hat{\alpha}$:

$$\hat{\alpha} = 1 + n \left(\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \right)^{-1}$$

Существует отдельная формула для дискретного случая.

Примеры

Что распределено согласно степенному закону?

1. Размеры лунных кратеров.
2. Частота употребления слов в большинстве языков.
3. Распространённость фамилий.
4. Масштабы аварий в энергосистемах.
5. Число уголовных обвинений на одного преступника.
6. Количество извержений вулканов.

Если $alpha \leq 2$

Существуют степенные распределения со свойством безмасштабности, среди которых, например, интенсивность солнечных вспышек, для которых сохраняется вероятность появления более мощного наблюдения, чем все предыдущие.

Модель Эрдёша-Реньи

Пуассоновский случайный граф

Пусть дан граф $G(N, E)$, принадлежащий ансамблю графов со множеством вершин N, в котором каждое ребро $e \in E$ принадлежит множеству рёбер полносвязного графа на N с вероятностью p. Вероятность появления конкретного графа G при заданном p выражается формулой:

$$P^G = p^m(1-p)^{C_n^2 - m},$$

где m — это количество рёбер в G, а $C_n^k = \frac{n!}{(n-k)!k!}$ — количество сочетаний из n по k.

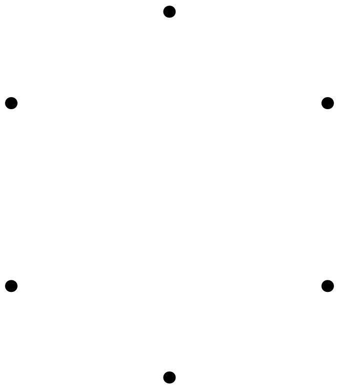
Характеристики

Ключевые оценки модели Эрдеша-Реньи

1. Суммарная степень всех вершин $2m$.
2. Среднее количество рёбер $m = C_n^2 p = \frac{n(n-1)}{2} p$.
3. Средняя степень $c = \frac{2m}{n} = \frac{2n(n-1)p}{2n} = (n-1)p$.
4. Распределение степеней $p(x) = \frac{c^x}{x!} e^{-c}$ — плотность распределения пуассона.
5. Коэффициент кластеризации равен просто p или $\frac{c}{n-1}$, но $\lim\limits_{n-to-infty} \frac{c}{n-1} = 0$, что явно говорит о определённо невысокой правдоподобности модели.

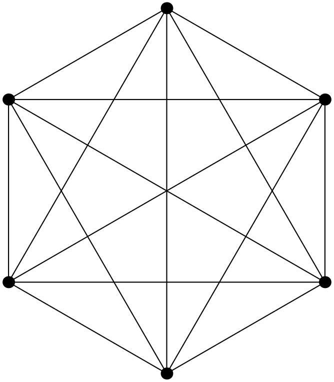
$p = 0$

Отдельные вершины графа



$p = 1$

Полносвязный граф



Фазовый переход

И появление гигантского компонента

Определение

Гигантский компонент графа — это компонент, размер которого сравним с количеством вершин в графе.

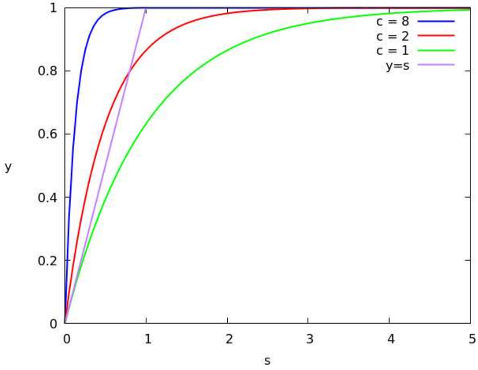
При $p = 0$ гигантский компонент отсутствует, а при $p = 1$ гигантский компонент занимает весь граф. Значение p, в котором в графе появляется гигантский компонент, называют точкой фазового перехода.
Если через s обозначить размер гигантского компонента, то он существует, если следующее уравнение имеет решение:

$$s = 1 - e^{-cs},$$

где c — это средняя степень вершин в графе, выражаемая через p.

Фазовый переход

Точка появления гигантской компоненты



Модель Уоттса-Строгаца

Повысим коэффициент кластеризации

1. Возьмём решётку, в которой каждый узел связан с c соседями ($c = 4$ — квадратная решётка).
2. Выбираем случайное ребро из решётки и заменяем его на случайное другое ребро.
3. Продолжая процесс, мы будем переходить от решётки к случайному графу. В процессе замен кластеризация и диаметр будут уменьшаться.

Интересное наблюдение!

Незначительное количество замен уже приводит к появлению эффекта маленького мира.

Блочная модель

И свойство ассортативности

При назначение узлам некоторого свойства

Сеть ассортативна, если узлы чаще связываются с узлами, имеющими то же значение свойства. Если чаще не связываются, то сеть дизассортативна.

Блочная модель — разновидность модели Эрдёша-Реньи, в которой все вершины распределены по группам, и задана матрица вероятностей появления ребра между вершинами разных групп.
Использование блочной модели позволяет моделировать сети со свойством ассортативности.

Спасибо за внимание!

Всё о курсе: https://marigostra.ru/materials/networks.html
E-mail: msp@luwrain.org
Канал в Телеграм: https://t.me/MarigostraRu