Виды и свойства попарной зависимости длин интервалов между последовательными эпизодами в пуассоновской и гамма–пуассоновской моделях поведения

Столярова Валерия Фуатовна

лаборатория теоретических и междисциплинарных проблем информатики, СПИИРАН, Санкт-Петербург

ISYT, 30.06.2017

Моделирование поведения индивида.

- Представляет собой последовательность событий на временной оси, каждое из которых связано с тем или иным риском.
- В ряде областей эпидемиологии, компьютерной безопасности существует задача оценки интенсивности поведения индивида.
- Доступны данные лишь о последних эпизодах поведения и/или рекордных интервалах между эпизодами.
- Математическая модель: пуассоновский процесс (**пуассоновская модель поведения**) или пуассоновский процесс со случайной интенсивностью (**гамма–пуассоновская модель поведения**).

Вопрос

Какими способами можно оценить интенсивность поведения?

Методы оценки интенсивности поведения.

- Байесовские сети доверия.
- Прямая оценка методом максимального правдоподобия.

Задача

Определить характер зависимости интервалов между эпизодами поведения в пуассоновской и гамма–пуассоновской моделях поведения.

Копулы.

Определение

*n–**копула** представляет собой ограниченное на единичный куб I^n совместное распределение вероятности n случайных величин с равномерным распределнием вероятности на отрезке I.*

- отметим, что для любой непрерывной случайной величины X с распределением вероятности $F(x)$, случайная величина $U = F(X)$ будет иметь равномерное распределение;
- теорема Склара устанавливает существование такой функции C, что совместная функция распределения n случайных величин $X_1, X_2, \ldots, X_n$ имеет вид:

$$H(x_1, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)).$$

Свойства копул.

- Монотонное преобразование величин, связанных копулой, **либо не изменяет копулу вовсе, либо изменяет её предсказуемым образом**.
- При моделировании зависимости копулы позволяют разделить оценку параметров маргинальных распределений и оценку параметров зависимости (структуры).
- Копулы являются липшицевыми функциями: любая n-мерная копула C удовлетворяет неравенству

$$|C(s_1, \dots, s_n) - C(t_1, \dots, t_n)| \leq |s_1 - t_1| + \dots + |s_n - t_n|.$$

Система интервалов между последовательными эпизодами поведения.

- τ_1 и τ_2 — интервалы между последовательными эпизодами поведения. В пуассоновской и гамма–пуассоновской моделях τ_1 и τ_2 имеют экспоненциальное распределение $F(x) = P[\tau_1 < x] = 1 - e^{-\lambda x}$ с параметром $\lambda > 0$.
- В рамках пуассоновской модели поведения, интервалы τ_1 и τ_2 являются независимыми.
- В рамках гамма–пуассоновской модели поведения

$$P[\tau_1 < x, \tau_2 < y|\lambda] = P[\tau_1 < x|\lambda]P[\tau_2 < x|\lambda].$$

- Обозначим копулу, описывающую взаимосвязь между τ_1 и τ_2, как

$$C_{12}(u, v) = P[F(\tau_1) < u, F(\tau_2) < v].$$

Копула системы интервалов между последовательными эпизодами поведения в пуассоновской модели.

$$C_{12}(u, v) = P[F(\tau_1) < u, F(\tau_2) < v] =$$
$$= P[F(\tau_1) < u] \cdot P[F(\tau_2) < v] = uv = \Pi.$$

Π — *копула независимости* или *копула произведения*.

Копула системы интервалов между последовательными эпизодами поведения в гамма–пуассоновской модели.

$$P[\tau_1 > x, \tau_2 > y] = \frac{\beta^{\alpha+1}}{(x + y + \beta)^{\alpha+1}}.$$

Пусть $\beta = 1$. Тогда полученное распределение вероятности является *двумерным распределением Парето II типа*.

$$\hat{C}_{12}(u, v) = (u^{-1/\theta} + v^{-1/\theta} - 1)^{-\theta}.$$

$\hat{C}_{12}(u, v)$ является *копулой Клейтона*.

Копула Клейтона. Свойства процесса поведения.

1. Параметр копулы может принимать значения $[-1,0)\cup(0,\infty)$, в случае $\theta=0$, $C_{12}(u,v)=\Pi$, где Π — копула независимости.
2. Копула Клейтона является архимедовой копулой с производящей (generating) функцией $\phi(t)=t^{-1/\theta}-1, t>0$.
3. Кроме того, для строгого подкласса копул Клейтона возможно обобщение на многомерный случай: пусть $\mathbf{u}=(u_1,\ldots,u_n), n\geq 2$, тогда

$$C_\theta^n(\mathbf{u})=(u_1^{-1/\theta}+u_2^{-1/\theta}+\ldots+u_n^{-1/\theta}-n+1)^\theta, \theta>0, n\geq 2.$$

Копула Клейтона. Свойства процесса поведения.

1. Переменные, связанные копулой Клейтона, обладают свойством PQD (Positive Quadrant Dependence):

$$P[\tau_1 \geq x, \tau_2 \geq y] > P[\tau_1 \geq x]P[\tau_2 \geq x].$$

   Это свойство говорит о том, что вероятность того, что τ_1 и τ_2 одновременно принимают большие значения выше, чем если бы они были независимы.

2. Для переменных, обладающих свойством PQD выполнено соотношение между коэффициентами ранговой корреляции τ-Кендалла и ρ-Спирмена: $3\tau_{XY} \geq \rho_{XY} \geq 0$, и $\mathrm{cov}(X, Y) \geq 0$.

Заключение.

- Был выведен вид копулы описывающий зависимость длин интервалов в рамках пуассоновской и гамма-пуассоновской моделей поведения.
- Копула Клейтона обладает рядом важных свойств: возможностью обобщения на многомерный случай, наличием алгоритмов для симуляции данных связанных такой копулой.
- Копула Клейтона может использоваться вкупе с другими распределениями длин интервалов (не только экспоненциальными), что является основой для создания новых моделей поведения.
- Копулы используются в вероятностных графических моделях, не требующих дискретизации переменных.

Спасибо за внимание!