Chị thương ơi

Chị ơi em nghĩ được cách tính bách phân vị đa biến rồi, rất tuyệt vời nhưng không phức tạp đâu

Chị sẽ thích lắm

Ví dụ nhe, chị tính bách phân vị cho 2 số đo: ĐKLĐ, CVB

Giả sử ĐKLĐ phân thành m khoảng *d1, d2, d3,…, dm* và CVB phân thành n khoảng *c1, c2, c3,…, cn*. Lưu ý: chị nên chuẩn hóa về [0… 1] và định khoảng sao cho *m = n* nhưng pp mình tính sẽ tổng quát.

Chúng mình sẽ đo trọng lượng thai (TLT) khi ĐKLĐ rơi vào khoảng *d1 =* [*60 – 70*] và CVB rơi vào khoảng *c2=* [*200 – 240*]. Mình sẽ không bàn đến trường hợp quá đơn giản khi đo có xác định cùng rơi vào hai khoảng vì việc này không thích hợp khi có nhiều số đo ( > 2). Các bước thực hiện như sau:

1. Đo TLT rơi vào *d1*, giả sử *tlt1* = {*1000, 1100, 1200*}. Đo TLT rơi vào *c2*, giả sử *tlt2* = {*1100, 1200, 1300, 1400, 1500*}.
2. Lấy phần hội của *tlt1* và *tlt2*: *tlt* = *tlt1* *U* *tlt2* = {*1000, 1100, 1200, 1100, 1200, 1300, 1400, 1500*}
3. Chuẩn hóa *tlt* về phân bố chuẩn tắc theo công thức: (*X – μ*) / (*σ / sqrt*(*n*)) với *μ* và *σ* lần lượt là trung bình mẫu và phương sai mẫu của *tlt*, *n* là cỡ mẫu (ở đâu là *8*), *sqrt* là căn bậc hai
4. Chị tra bảng phân bố chuẩn sẽ xác định được bách phân vị của *tlt* đã chuẩn tắc.
5. Tính lại số bách phân theo công thức: *Xσ/ sqrt*(*n*) + *μ*

Chị có biết vì sao mình phải chia khoảng?

* Số tổ hợp sẽ ít đi rất nhiều, thậm chí trường hợp số thực vẫn tính được. Chị lập bảng biểu dễ dàng hơn
* Việc xác định khoảng sẽ thiết thực hơn khi chị thông tin với bệnh nhân hay khi bảo vệ luận văn. Mình chỉ cần ước lượng khoảng.

Tinh yếu thuật giải:

* Phân khúc (định khoảng)
* Lấy phần hội (không phải phần giao đâu nhe). Chị lưu ý điểm này, đúng cho cả trường hợp một biến, nhưng mình đã định nghĩa lại bách phân vị. Em sẽ giải thích bằng lời.

Bước *3, 4, 5* với giả định tập *tlt* theo phân bố chuẩn, chúng mình sẽ cải tiến tính bách phân vị thứ *k* cho trường hợp tổng quát phân bố bất kỳ.

1. Sắp xếp và rút gọn tập *tlt* thành tập *measure =* {*1000, 1200, 1300, 1400, 1500*}
2. Cho *i* chạy từ phần tử thứ nhất *measure*[*1*] = *1000* đến phần tử cuối *measure*[*5*]=*1500*

Tính tổng xác suất với *Pj* là xác suất của phần tử thứ *j* (*measure*[*j*]). Ví dụ: xác suất của phần tử thứ *2* (*measure*[*2*] = *1200*) là *P2 = 2 / 8 = 0.25.* Tương tự *P1 = 1 / 8 = 0.125*

1. Nếu *s > k* thì *i* chính là bách phân vị thứ *k*. Ví dụ: nếu *k = 0.3, s2 = P1 + P2 = 0.25 + 0.125 = 0.375 > k*, vậy bách phân vị *3th* là *1200*.

Chị ơi, em nghĩ lấy phần hội chưa chính xác. Thuật toán sẽ cải tiến theo hướng khác, với giả định “*các độ đo độc lập từng đôi một tạo thành không gian vector*”, khi đó xác suất TLT bằng tích các xác suất rơi vào từng khoảng. Ví dụ xác suất TLT trong trường hợp trên bằng tích của xác suất rơi vào *d1* và xác suất rơi vào *c2*. Các bước được cải biến như sau:

1. Đo TLT rơi vào *d1*, giả sử *tlt1* = {*1000, 1000, 1000*}. Đo TLT rơi vào *c2*, giả sử *tlt2* = {*1000, 1200, 1200, 1200*}.
2. Sắp xếp và rút gọn tập *tlt* thành tập *measure =* {*1000, 1200*}
3. Cho *i* chạy từ phần tử thứ nhất *measure*[*1*] = *1000* đến phần tử cuối *measure*[*2*]=*1200.* Tính xác suất tích *pi =* với *ai* và *bi* lần lượt là xác suất để *measure*[*i*] rơi vào *d1* và *measure*[*i*] rơi vào *c2*. Ví dụ, *a1 = 3/3 , b1 = 1/4, p2 = 1\*1/4 = 0*.25
4. Nếu *s > k* thì *i* chính là bách phân vị thứ *k*. Ví dụ: nếu *k = 0.2, p2 = 0.25 + 0\*3/4 = 2.25 > k*, vậy bách phân vị *2th* là *1200*.

Tinh yếu thuật giải:

* Phân khúc (định khoảng)
* Các độ đo độc lập từng đôi một, xác suất tại một phân vị bằng tích các xác suất độ đo thành phần.

Và dễ thương nhất