

الفصل الرابع

فهم شروط إحلال التوزيع

استقراء ما بعد البيانات

تعمل ماريا في مستشفى تعليمي كبير وتتحمل مسؤولية طلب الإمدادات، وتواجه اليوم موقف صعب: كما أوضحت لصديقتها "آنا" "أنا بحاجة إلى بيانات ثلاث أعوام ولكن ليس لدي سوى عامين". فطلبت منها آنا أن توضح قصدها، وبدأت ماريا في سرد مشاكلها التي وقعت فيها خلال الأيام القليلة الماضية.

"على مدى العامين الماضيين كان لدينا اختبار تشخيصي لنوع معين من مرض الشعب الهوائية الذي يصيب الرضع وقد تم تسجيل المرضى الذين يعانون من هذه الحالة، وكان هناك دواء فعال يعالج هذا المرض. ولكن كان هناك نقص في المعروض بالإضافة إلى ارتفاع تكلفته. والأسوأ من ذلك، أنه يحتوي على بعض المكونات غير المستقرة لذلك كان يجب استخدامه في غضون الستة أسابيع الأولى بعد التصنيع. وكانت الطلبات العاجلة تستغرق أسبوعاً حتى تصل من منشأة التصنيع بسويسرا، لذلك كان يجب علينا الحفاظ على المخزون عن ما لا يقل عن أسبوع، ويجب علي الآن تحديد الكمية المطلوبة.

وفي هذا النوع من الحالات، تشير قواعدا المعتادة إلى أنه ينبغي نفاذ المخزون أكثر من مرة كل ثلاثة سنوات. أستطيع أن استنتج كم كنا سنحتاج في كل من الأسابيع الـ 104 الماضية، ولكن هذه ليست فترة طويلة بما فيه الكفاية "

أنهت "أنا" للتو دورة إدارة المخاطر كمرحلة من ماجستير في إدارة الأعمال وتتساءل عما إذا كان هناك طريقة لاستخدام بعض ما درسته لمساعدة صديقتها. وقالت " لديك بيانات 104 أسبوع ومشكلتك هي كيفية تقدير أعلى رقم إذا تم تمديد هذه لسلسلة إلى ثلاث سنوات، أي $3 \times 52 = 156$ أسبوع من البيانات. "

ردت ماريا "هذا هو بالضبط". "دعينا نفكر في ذلك، حتى في حال كان هناك ثلاث سنوات من البيانات فإنها لن تكون كافية - لمجرد أن شيئاً لم يحدث على مدى السنوات الثلاث الماضية لا يعني أنه لن يحدث على مدى السنوات الثلاث المقبلة." هل فكرت في وضع نموذج للاستخدام الأسبوعي مع التوزيع العادي، أو ربما نوعاً آخر من التوزيع؟ "

ردت ماريا "كانت لدي نفس الفكرة، ولكن نظرت إلى الأرقام وكان هناك المزيد من الأسابيع يشهد فيها معدل الطلب ارتفاعاً مما يضعف احتمالات ما إذا كان سيحدثي التوزيع الطبيعي نفعاً حقاً. يبدو أن ما حدث يرجع إلى كثير من العوامل التي لا أعرف عنها شيئاً. أشياء مثل الطقس، وعدد من حالات التهابات الصدر. لا أعتقد أنني أستطيع استخدام نموذج لوضع توقعات، هناك الكثير من حالات عدم اليقين، لذلك نحن أمام بيانات غير كافية."

قالت أنا "لذا، أنتي بالحاجة إلى استخدام البيانات المتاحة لديك، ولكن استقراء قيم عالية لم تحدث بعد، وفي الوقت نفسه لا تريدين أن تضعي أي افتراضات حول هذا النوع من التوزيع".

"هذا صحيح: بدون بيانات كافية وبدون نموذج محدد للتوزيع يبدو من المستحيل تقدير العدد الذي أحتاج إليه".

قالت أنا: "أنا لست متأكدة من ذلك، حتى من دون افتراض أن الأرقام الأسبوعية

تتطابق مع توزيع معين، فإنها بالتأكد لا يمكن أن تكون شاذة للغاية. أي نوع من الانتظام يجب أن يوفر لكي دلائل على استقرار المشكلة. قد يكون من الصعب استخراج معلومات حول هذا "الذيل (المنحنى) العلوي" للتوزيع من البيانات المتاحة لديك، ولكن ربما ليس مستحيلا. سمعت أستاذي يتحدث عن نظرية القيمة القصوى وربما يمكن أن تساعد بطريقة ما. "

1-4 التوزيع ذو القيم القصوى (غير المحدود)

1-1-4 تعريف مؤشر المنحنى (شروط الإحلل)

إن التحدي المتمثل في استخدام التقنيات الكمية لقياس المخاطر هو أنه يجبرنا على الانتباه إلى ذيول (منحنى) التوزيع. هذا هو بالضبط مدى الصعوبة في التوقع والتقدير: نادرا ما نرى ما يكفي من الأحداث لوضع استنتاجات محددة حول التوزيعات الأساسية. في هذا الفصل سوف ننظر بمزيد من التفصيل في بعض الأدوات للتعامل مع ذيول التوزيعات.

عند التعامل مع المخاطر نهتم كثير من الأحيان بالمتغيرات العشوائية التي لديها احتمالية (على الأقل نظريا) من القيم الكبيرة بلا حدود. مثل التوزيع الطبيعي، الذي له متغيرات عشوائية ليس لها نطاق محدود. ومن الواضح أنه من الناحية العملية سيكون هناك في الغالب حد للتوزيع: فعلى سبيل المثال، إذا تعاملنا مع توزيع الخسارة، ستكون هناك خسارة قصوى تحددها قدرة شركتنا على تجنب الإفلاس. ولكن عادة ما يكون أكثر تكشفا عن نموذج الخسائر كما لو لم يكن هناك حد أقصى. في هذا السياق من المنطقي أن ننظر إلى شكل ذيل التوزيع الذي ليس له نهاية.

التوزيع الطبيعي يعطي نقطة طبيعية للمقارنة بالتوزيعات الأخرى. ويسمى هذا التوزيع بـ "التوزيع غير المحدود" إذا كان لديه المزيد من الوزن في المنحنى بأكثر من التوزيع الطبيعي. ولكن علينا أن نتوقف ونفكر فيما قد نعنيه بهذه الجملة. لأنه يمكن تعيين الانحراف المعياري من العادي إلى ما نرغب به، فمن الممكن دائما للعثور على التوزيع الطبيعي الذي يعني

صفر ولكن هناك احتمال كبير لأن يكون أكبر من قيمة معينة. نحن بحاجة لجعل الانحراف المعياري كبير بما فيه الكفاية. لذلك، لفهم التوزيع غير المحدود، نحن بحاجة للتفكير في سلوك توزيعها كمجموعة كاملة من القيم وليس فقط في نقطة واحدة.

طريقة واحدة لفهم سلوك منحنى التوزيع هو أن ننظر في دالة التوزيع التراكمي لنهج التوزيع وهناك طريقة جيدة للقيام بذلك وهي النظر في الناتج $(1 - F(x))X^K$ لبعض K . المصطلح الأول $1 - F(x)$ أقرب إلى الصفر مع زيادة x ، في حين أن المدى الثاني X^K سوف يكون أكبر وأكبر، لذلك يمكننا أن نسأل أي من هذين سيكون الأصح؟ هل سيتقرب الناتج من الصفر أو الاتجاه إلى ما لا نهاية؟

يمكننا أن نحسن أننا نجعل k أكبر، وسيكون في مرحلة ما k_0 حينها XK تكون الأثقل وزنًا. بالنسبة إلى $k < k_0$ فإن الناتج سوف يقترب من الصفر، وبالنسبة إلى $k > k_0$ فإن الناتج سيتجه للانهاية. إذا كان ذيل التوزيع ذو قيم كبيرة، سيكون هناك فرص كبيرة للقيم المرتفعة وهذا يعني أن $F(X)$ تقترب من الواحد ببطء. لذلك لن نكون قادرين على مضاعفة مثل هذه القيمة الكبيرة من X^K والاستمرار في الحصول على ناتج يقترب من الصفر. وبالتالي، يرتبط المنحى ذو القيم المرتفعة بقيمة منخفضة من K_0 .

عندما يمكننا تحديد قيمة k_0 بهذه الطريقة، مع $(1-F(x))x^k$ باتجاهها إلى الصفر أو اللانهاية وفقا لما إذا كان k إما أقل أو أعلى من k_0 ، ثم نقول أن التوزيع له مؤشر المنحنى (الذيل) من k_0 . لذلك، على سبيل المثال، إذا كان مؤشر المنحنى (الذيل) يساوي 2 ويساوي العبارة $1 - F(x)$ تتجه إلى الصفر بنفس الطريقة مثل $1/x^2 = x^{-2}$. ولكن التعريف الذي قدمناه، من حيث نقطة فاصلة بين نظامين، أكثر دقة.

مثال 1-4 التوزيع ذو القيم القصوى (غير المحدود)

متغير عشوائي لديه توزيع مع دالة الكثافة

$$f(x) = \frac{1}{(2 + |x|)^2}$$

هذا متماثل حول 0 لأن قيم f لا تعتمد على علامة x . وتتمثل دالة الكثافة في الرسم البياني 1-4. يمكننا التحقق من أن هذا هو توزيع من خلال إظهار أن النهج متكامل 1 كما $x \rightarrow \infty$. للقيام بذلك نلاحظ أن $z > 0$

$$\int_0^z f(x)dx = \int_0^z \frac{1}{(2+x)^2} dx = \left[\frac{-1}{2+x} \right]_0^z = \frac{1}{2} - \frac{1}{2+z},$$

التي تقترب بوضوح من $1/2$ حيث $z \rightarrow \infty$ ، وبالتالي فإن المنطقة تحت المنحنى كلها ستكون 1 كما نرغب. لتحديد دالة التوزيع التراكمي يمكننا استخدام التماثل لمعرفة ذلك

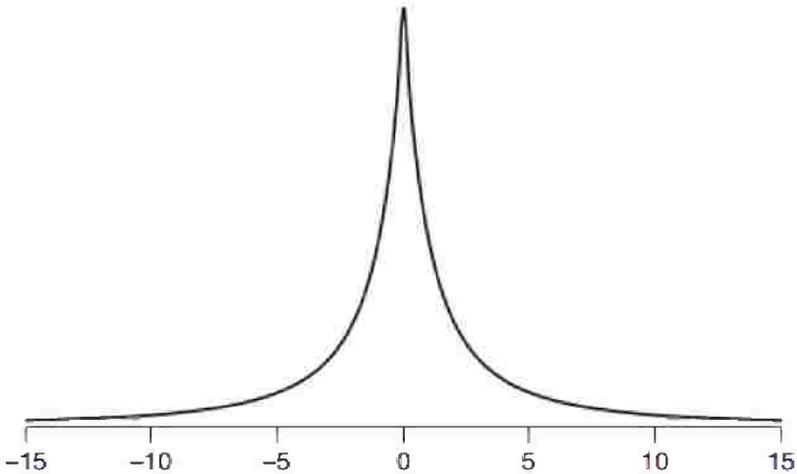
$$F(0) = \int_{-\infty}^0 f(x)dx = \frac{1}{2}$$

$z > 0$,

كما

$$F(z) = \int_{-\infty}^0 f(x)dx + \int_0^z f(x)dx = 1 - \frac{1}{2+z}.$$

وبالتالي، $1 - F(z) = (2+z)^{-1}$ وهذا التوزيع له مؤشر المنحنى (الذيل) عند 1. هذه هي حالة شاذة من التوزيع ذو القيم القصوى.



الشكل 1-4: مثال لدالة الكثافة مع مؤشر المنحنى (الذيل) المساوي لـ 1.

من المهم أن ندرك أن ليس لكل التوزيعات تباين وانحراف معياري واضح المعالم. ويُعرّف التباين لمتغير عشوائي X بأنه $E[(X - E(X))^2]$ وإذا كان التوزيع له منحني ذو قيم قصوى فإن هذا التوقع قد يكون لانهائي. وفي هذه الحالة يكون الحديث عن الانحراف المعياري للمتغير العشوائي بلا معنى. كلما كان مؤشر المنحني (الذيل) أصغر، كلما كان التوليف من التوزيع أكثر احتمالاً، والأرجح هو ألا يكون هناك تباين. والنتيجة الرئيسة هنا هي أنه إذا كان مؤشر المنحني (الذيل) أقل من k إذاً $E(X^k)$ سيكون لانهائي (نبين أسباب صحة هذا الأمر في القسم التالي).

هناك مجموعة كاملة من التوزيعات يكون فيها مؤشر المنحني (الذيل) لانهائي بوجود كافة البيانات. التوزيعات، مثل التوزيعات العادية، حينها تكون الاحتمالية في المنحني، $1 - F(x)$ ، تقترب من 0 أسرع من أي x . على الجانب الآخر، توزيع مثل الشكل 1-4 الذي يظهر مشكلة تحديد المتوسط. ومن الواضح أن التوزيع متماثل حول الصفر، وبالتالي يجب أن يكون متوسط يساوي الصفر، ولكن إذا نظرنا فقط في ذيل التوزيع سينصب اهتمامنا على تكامل المعادلة $\int_u^{\infty} xf(x)dx$ لبعض القيم u ، وتبين أن هذا التكامل له قيمة لانهائية.

2-1-4 توقع مؤشر المنحني (الذيل)

بالنظر إلى مجموعة من نقاط البيانات قد نرغب في تقدير مدى سرعة منحني (ذيل) التوزيع في الاتجاه إلى الصفر لمعرفة ما إذا كنا نتعامل مع توزيع ذو حدود قصوى غير محدودة. على سبيل المثال، هناك سؤال يجب الإجابة عليه إذا احتجنا إلى توقع النقص المتوقع من البيانات. في وقت لاحق من هذا الفصل سوف نناقش بشكل أكثر تفصيلاً للطريقة التي يمكن أن يتم بها تقدير قياس المخاطر باستخدام نظرية القيمة القصوى، ولكن هنا نريد أن نعطي نهجاً أكثر أولية لتقدير مؤشرات الذيل.

دالة التوزيع التراكمي، لنفترض أن التوزيع له مؤشر ذيل α ، ثم، بالنسبة للقيم الكبيرة من x ، نتوقع أن تعطى دالة توزيع تراكمي تقريبا من خلال

$$F(x) = 1 - kx^{-\alpha}$$

لبعض K الثابتة. إذا كان صحيحًا إذًا

$$\log(1 - F(x)) = \log(kx^{-\alpha}) = \log(k) - \alpha \log(x).$$

بما أن $1 - F(x)$ بين الـ 0 و 1 وسجل الأرقام (log) أقل من 1 سلبى، سيكون

مساعد في ضرب هذه المعادلة في -1 للحصول على

$$-\log(1 - F(x)) = -\log(k) + \alpha \log(x).$$

هذا يدل على أنه إذا حددنا السجل $-\log(1 - F(x))$ ضد سجل $\log(x)$ إذًا يجب

أن نحصل على خط مستقيم مع ميل α .

وبالنظر إلى مجموعة من نقاط البيانات فإنه من السهل تقدير $1 - F(x)$ من خلال

النظر في نسبة النقاط في العينة فوق القيمة x . وبالتالي، فإن الإجراء هو تحديد نواقص

السجل \log لهذا التوقع ضد سجل $\log x$ لتقدير مؤشر المنحنى (الذيل).

مثال 2-4 سلوك المنحنى (الذيل) لأسعار الصرف في سوق العملات

لتوضيح ذلك، يوضح الشكل 2-4 مخطط بياني من هذا النموذج لسعر صرف الجنيه

الإسترليني والدولار الأمريكي لفترة 500 يوم تبدأ في مايو 2010. وتظهر البيانات التغير

النسبي اليومي في سعر الإغلاق. ننظر هنا في الخسائر من يوم لآخر إذا تم شراء الجنيه

الإسترليني (الجنيه ينخفض أمام الدولار)، لذلك يعطي هذا 499 نقطة بيانات مختلفة.

خلال هذه الفترة كانت أكبر خسارة 1.74٪ في 13 مايو 2010. يتم رسم ذلك في هيئة 'x'

قيمة سجل $\log(1.74) = 0.55$ (باستخدام سجلات طبيعية). هذه النقطة لديها 498

نقطة دون ذلك من تعدد خسائر ولذا فإننا نقدر قيمة $F(x)$ لتكون $499/498$ وبالتالي فإن

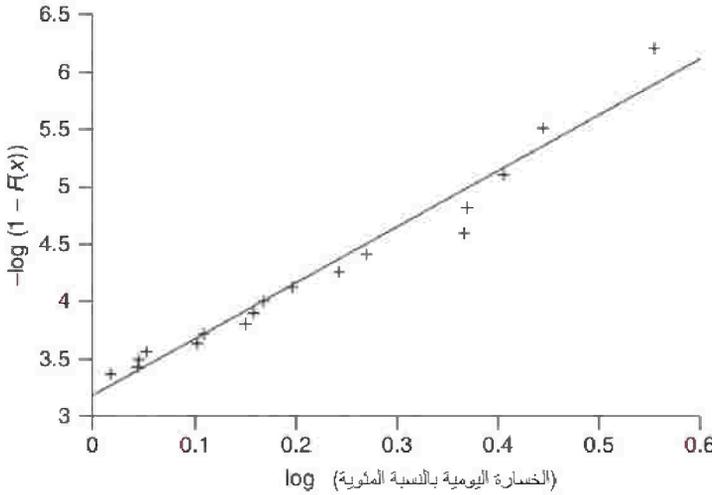
قيمة y في المخطط تعطى من خلال

$$-\log(1 - F(x)) = -\log(1/499) = 6.21.$$

يبين الشكل حركة السعر على مدار 17 يومًا والتي تخطت الخسائر فيها أكثر من 1٪.

الخط المستقيم الواصل بتلك النقاط له ميل 4.9. لذلك، من هذه البيانات وحدها،

نستطيع توقع مؤشر الذيل بنحو 5، وهناك بعض الأدلة على السلوك المنحنى غير المحدود.



الشكل 2-4: بيانات المنحنى (الذيل) للخسارة اليومية بالنسبة المئوية في سعر صرف الإسترليني/دولار

عندما تكون المنحنيات غير محدودة فسوف يكون مؤشر الذيل لا نهائي. بطبيعة الحال، سينتج عن ذلك ميل قوي ومن خلال رسم مخطط باستخدام توزيع طبيعي سيكون هناك ميل أكثر من المبين في الشكل 2-4. بالرغم من ذلك، بما أن لدينا عدد غير محدود من الأمثلة فإنه من الناحية العملية من المستحيل معرفة سلوك المنحنى تحديداً. وكلما زادت أعداد الأمثلة كلما كانت قدرتنا على التوقع أفضل، إلا أن تعريف الذيل هو شيء يحدث في القيم القصوى ولن يكون لدينا أبداً بيانات كافية للتأكد من هذا الأمر.

مثال 3-4 النظر في سلوك الذيل للتوزيع الطبيعي

يمكننا محاكاة البيانات ببساطة من التوزيع الطبيعي. يظهر الشكل 3-4 نتائج مثالين مختلفين للتوزيع الطبيعي بمتوسط 0 وانحراف معياري 1. ويوجد على يسار المخطط مثال يحتوي على 500 نقطة من البيانات. ويظهر الشكل سجل $-\log(1 - F(x))$ على المخطط مقابل السجل $\log(x)$ لأعلى 20 نقطة؛ وأعلى نقطة في هذا المثال هي 2.859. والخط المستقيم الواصل بين هذه النقاط له ميل يساوي 5.38. لاحظ أن هذا المثال المختلف قد ينتج عنه

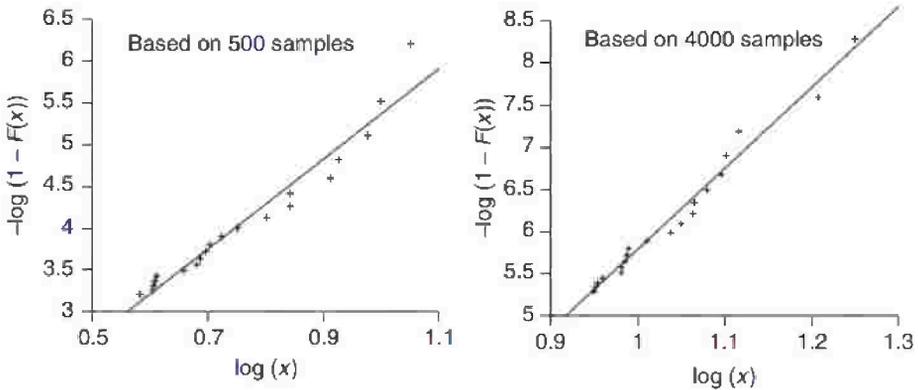
توقعات مختلفة تمامًا.

وعلى يمين الشكل، يوجد نموذج جديد يحتوي على 400 نقطة من البيانات، وأيضًا يظهر الشكل أعلى 20 نقطة. وفي هذه المرة فإن أعلى نقطة 3.48 وتتوافق مع قيمة متوقعة لـ

$$F(X) \downarrow 4000/3999 \text{ تؤدي إلى قيمة } Y \text{ في المخطط}$$

$$- \log(1 - F(x)) = - \log(1/4000) = 8.29.$$

الخط المستقيم الواصل بين النقاط على اليمين له ميل 9.57 وبالتالي أعلى من الرقم الحاصل من النموذج الأصغر. يمكننا رؤية ذلك من خلال استخدام مثال أكبر، تتوافق الـ 20 نقطة مع جزء من التوزيع أبعد من المنحنى. هذا يعطي لهذه الطريقة فرصة أكبر في اختيار منحنى ذو قيم أصغر ويكون لمؤشر الذيل توقعات أكبر. ولكن يظهر هذا المثال أيضًا صعوبة التأكد من مؤشر الذيل على أساس النموذج الذي يحتوي على 100 نقطة. ويوضح بشكل عملي أن دلائل استخدام التوزيعات غير المحدودة لأسعار الصرف المستخدمة في المثال 4-2 ضعيفة جدًا.



الشكل 3-4: توقعين مختلفين لمؤشر الذيل (المنحنى) باستخدام أمثلة من التوزيع الطبيعي.

3-1-4* المزيد من التفاصيل حول مؤشر الذيل (المنحنى)

افترضنا في تعريفنا لمؤشر الذيل (المنحنى) أن هناك قيمة مفردة لـ K_0 التي هي بين النطاق الذي فيه $(1 - F(x))x^k$ تتجه إلى ما لا نهاية وفي النطاق الذي تتجه فيه إلى الصفر. ولكن ماذا عن النطاق الذي تكون فيه $(1 - F(x))x^k$ لا تقترب من الصفر ولا إلى ما لا نهاية؟ في هذه الحالة لن يكون هناك تعريف محدد لمؤشر الذيل. في حقيقة الأمر، يكون ذلك غير محتمل الحدوث، ولكن لا يمكننا تجاهل تلك الاحتمالية. سنرى ما سبب ذلك.

سنبدأ بأن نكون دقيقين في ما تعنيه الدالة $G(x)$ التي لا تتجه إلى الصفر أو إلى ما لا نهاية. عدم الاتجاه إلى ما لا نهاية يعني أن بعض القيم M مثل $G(x) < M$ لسلسلة لا نهائية من القيم $X: \dots, x_2, x_1$. ولهذا لا يهم كبر حجم قيمة X حيث يمكننا دائمًا إيجاد قيم أكبر يكون فيها $G(x) < M$. هذا هو القيمة المنطقية السلبية لاتجاه $G(x)$ إلى ما لا نهاية، وهو مساوي للقول بأن كل $M, G(x)$ ستكون أعلى من M والاستقرار أعلاه. وفي الاتجاه الآخر، حينها لا تتجه $G(x)$ إلى الصفر فإن ذلك يعني أن بعض القيم لـ m مثل $G(x) > m$ لسلسلة لا نهائية من قيم $x: \dots, x_2, x_1$. الآن افترض أن هذا حدث لـ $(1 - F(x))x^k$

$$F(x) \text{ لكل قيم } K \text{ في نطاق } [k_0 - \delta, k_0 + \delta]. \text{ الآن إذا كان}$$

$$(1 - F(x_i))x_i^{k_0 + \delta} < M \text{ for } x_i \rightarrow \infty,$$

إذا

$$(1 - F(x_i))x_i^{k_0} < Mx_i^{-\delta}.$$

لاحظ أن الجانب الأيمن من هذا التباين يتجه إلى الصفر حيث أن x_i تصبح أكبر. ونفس الطريقة تنطبق على

$$(1 - F(x_i))x_i^{k_0 - \delta} > m \text{ for } x_i \rightarrow \infty,$$

ويكون لدينا

$$(1 - F(x_i))x_i^{k_0} > mx_i^{\delta}$$

الجزء الأيمن يتجه إلى ما لا نهاية حيث x_i تصبح أكبر. ولهذا، في نفس الوقت تكون سلسلة من قيم X_i يكون فيها $(1 - F(x_i))x_i^{k_0}$ تتجه إلى ما لا نهاية أو اتجاهها إلى الصفر. الحالة الوحيدة التي يمكن حدوث ذلك فيها هي في حال كان نتاج الدالة $(1 - F(x_i))x_i^{k_0}$

تتجه إلى الارتفاع أو الانخفاض وارتفاع القمم أكثر وأكثر واقتراب القيعان من الصفر وتقترب من الصفر أكثر وأكثر. يمكن أن نبني مثل هذه الدالة إذا حاولنا، ولكنها بالتأكيد ليست شيئاً محتمل الحدوث من الناحية العملية.

عندما نتحدث عن مؤشر الذيل لـ K لن يتوجب علينا تحديد ما إذا كان التعبير الحاسم $L(x) = (1 - F(x))xk$ نفسه يتجه إلى الصفر. ففي حقيقة الأمر، هناك احتمالية لكل من هذه الخيارات، ولكن الدالة $L(x)$ يجب أن تتجه إلى ما لا نهاية أو إلى الصفر بشكل سريع. وهناك بعض الحالات المطلوبة لهذا الأمر: يجب أن تكون $L(x)$ ذات متغير بطيء بمعنى:

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$$

لأي من القيم لـ $t > 0$. ولهذا، على سبيل المثال، إذا افترضنا أن $t = 2$ مضاعفة القيمة X لا يمكن أن تتم (في حدود القيمة الكبرى لـ X) وتبدو كتطبيق أي مضاعف غير الـ 1. لاحظ إنه إذا كان $L(x) = kx^\beta$ ، لنا $L(tx) = k(tx)^\beta$ ، وهذا يمكن أن يساوي فقط 1 (و L ذات متغير بطيء) في حال كان $\beta = 0$.

هناك بعد التعقيدات هنا لا نرغب في الخوض فيها، فالحالة $(1 - F(x))x^\alpha$ ذات متغير بطيء أقوى من الحالة التي يكون فيها الأس α يمثل نقطة الانقسام بين دالات تقترب من الصفر ودالات تقترب من اللانهاية. (يمكننا رؤية ذلك من مراقبة الفترة الزمنية للدالة $2 + \sin x$ التي لا تتغير بشكل بطيء ولكنها سوف تتأثر بـ $x\epsilon$ بأي قيمة صغيرة من ϵ). الوضع الذي يكون فيه تغير بطيء هو الوضع المطلوب لإثبات النتائج ذات القيمة القصوى التي سوف نوضحها لاحقاً، حتى وإن كانت الطريقة التي وضعنا بها مؤشر الذيل أبسط.

كما ذكرنا سابقاً، هناك علاقة وثيقة بين الأوقات التي يكون فيها التوزيع ومؤشرات الذيل متواجدين. نبدأ بالنظر في الأوضاع التي تتواجد فيها الدالة $E(X_2)$. عندما يكون لـ X دالة كثافة f و CDF F يمكننا كتابة

$$E(X^2) = \lim_{R \rightarrow \infty} \int_{-R}^R f(x)x^2 dx.$$

هنا قمنا بكتابة الحد العلوي والسفلي للتكامل كـ $-R$ و R (بدلاً من اللانهاية) لأن الصيغة التكاملية $-\infty$ إلى ∞ معرفة فقط عندما تكون حدود $R \rightarrow \infty$ موجودة، والتساؤل حول تواجدها أو لا هو محط اهتمامنا هنا. والآن، اختيار نقطة عشوائية u والتكامل مع النطاق الموضح أعلاه u (لاحظ أن في هذا النطاق $x \geq u$) يمكن أن نحصل على عدم مساواة

$$\int_u^R f(x)x^2 dx \geq \int_u^R f(x)u^2 dx = (F(R) - F(u))u^2.$$

السماح لـ R تتجه إلى ما لا نهاية يوضح أن $\lim_{R \rightarrow \infty} \int_u^R f(x)x^2 dx \geq (1 - F(u))u^2$.

الآن ضع في الاعتبار التوزيع مع مؤشر الذيل α أقل من 2. ثم $(1 - F(u))u^2$ تتجه إلى ما لا نهاية. ولهذا، لأي رقم كبير لـ M يمكن أن نختار u مع $(1 - F(u))u^2 > M$ وأيضاً لـ M . $\lim_{R \rightarrow \infty} \int_u^R f(x)x^2 dx > M$. التكامل للنطاق بأكمله $-R$ إلى R يجب أن يكون أكبر من ذلك، على سبيل المثال؟

$$\lim_{R \rightarrow \infty} \int_{-R}^R f(x)x^2 dx > M.$$

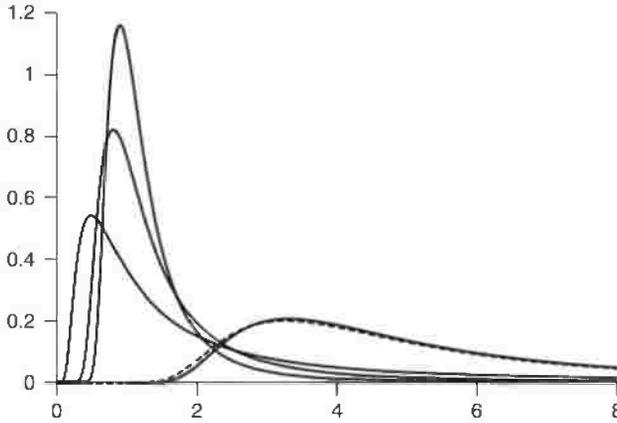
ولكن بما أن M يمكن اختيارها لأن تكون أي رقم نرغب به، سيحدد ذلك كـ $R \rightarrow \infty$ بعدم إمكانية حدوثها، على سبيل المثال. $E(X^2) = \infty$.

بشكل عام، يمكن استخدام هذه المناقشات في إظهار إذا كان التوزيع له مؤشر ذيل لـ α إذا الحالات التي يكون فيها $E(X^k)$ لن تقع لأي من $k > \alpha$.

2-4 تحديد التوزيع للقيم القصوى

في هذا القسم سوف ننظر في المشاكل التي تقع عندما يكون هناك قيم قصوى. اقترض أننا نريد توقع القيم القصوى لـ N لمتغيرات عشوائية مختلفة X_1, X_2, \dots, X_N . ولهم نفس التوزيع. على سبيل المثال، يمكن أن نهتم بالثروة القصوى لـ 100 فرد ضمن

عينة عشوائية، ولهذا فإن X_i تعني الثروة لدى i th (ترتيب الفرد). ونعرف أن في حال كانت X_i جميعها مستقلة فهناك N منهم وكل منهم له $F(x)$, CDF، وعليه فإن توزيع القيم القصوى التي نسميها بـ F_{max} ، لها $(F(x))^N$ CDF عندما تكون N كبيرة، تكون الذيل للتوزيع الأصلي فقط حيث أن أي قيمة لـ X حينها تكون $F(x)$ أقل من 1 سيكون لها $(F(x))^N$ رقم صغير جدًا. ويبدو هذا منطقيًا عندما نأخذ القيمة القصوى لرقم أكبر من الرسومات لمتغيرات عشوائية سنكون ملزمين لرؤية قيمة على الجانب الأيمن لمنحنى التوزيع كما هو الحال مع سلوك التوزيع في المنحنى الذي يكون الشيء الوحيد الذي يحسب.



الشكل 4.4 كثافات فريشيه لقيم مختلفة لـ α

الشكل 4.4: كثافات فريشيه لقيم مختلفة لـ α

سوف نوضح ذلك من خلال طريقة عرض التوزيع للقيم القصوى لـ N لعدد كبير من N سوف يعتمد فقط على مؤشر الذيل للتوزيع الأصلي. سنأخذ في اعتبارنا التوزيع ذو مؤشر الذيل α وبالأخص سنفترض أن F هي التوزيع الذي يكون فيه المنحنى $1 - F(x) = kx^{-\alpha}$ على سبيل المثال $F(x) = 1 - kx^{-\alpha}$ نحتاج إلى تعريف مضاعف محدد b_N والذي سوف يكون بمثابة مقياس.

$$b_N = (kN)^{1/\alpha}.$$

إذن

$$F_{\max}(b_N x) = (F(b_N x))^N = (1 - kb_N^{-\alpha} x^{-\alpha})^N.$$

لكن

$$b_N^{-\alpha} = ((kN)^{1/\alpha})^{-\alpha} = (kN)^{-1},$$

ولهذا

$$F_{\max}(b_N x) = \left(1 - \frac{x^{-\alpha}}{N}\right)^N. \quad 1-4$$

سنحتاج إلى استخدام العديد من e في هذا القسم وفي بعض الأحيان يكون استخدام رموز $\exp(x)$ أوضح للمتوسط ex . سوف نتقل بين هذين الاثني (وأيضاً جمعهم) بهدف تجنب المعادلات المطولة من المعدلات الأكثر طولاً لكي يكون التعبير عنها أبسط للقراءة. حقيقة واحدة عن استخدام الدالة الأسية هي أنها يمكن أن تكون بمثابة حد:

$$\exp(x) = \lim_{N \rightarrow \infty} (1 + x/N)^N, \quad 2-4$$

ولهذا، من خلال المعادلة (1-4) نرى أن عندما تكون N كبيرة يكون لدينا $F_{\max}(b_N x)$ قرابة $\exp(-x-a)$.

وقد أضحنا أنه إذا أخذنا القيم القصوى لـ 50 متغير عشوائي لمستقبل موزع يكون فيها $1 - F(x)$ تتجه إلى الصفر مثل $x^{-\alpha}$ إذا يكون التوزيع تقريباً نسخة مقيسة من واحد مع دالة التوزيع $\exp(-x-a)$ ، حيث يبقى الشكل كما هو ولكن المحور الأفقي يتم مضاعفته من خلال $b_{50} = (50k)^{1/\alpha}$.

ولهذا، مؤشر الذيل α هو الشيء الوحيد الذي يحدد شكل التوزيع والتوزيع للقيم القصوى التي لها دالة توزيع تراكمي والتي تكون $F(x) = \exp(-x-a)$ ولكن هذا يكون له قياس من خلال الكمية التي تعتمد على N والتوزيع المطبق.

التوزيع باستخدام دالة التوزيع التراكمي يكون من خلال $F(x) = \exp(-x-a)$ ويسمى بـ "Fréchet distribution" على اسم عالم الرياضيات الفرنسي البارز موريس

فريشيه. دالة التوزيع لـ فريشيه يتم الحصول عليها باستخدام قواعد "حساب التفاضل والتكامل:

$$(d/dx) \exp(g(x)) = g'(x) \exp(g(x))$$

حيث نكتب $g'(x)$ لمشتقة الدالة $g(x)$. وعليه تكون الكثافة

$$f(x) = \alpha x^{-\alpha-1} \exp(-x^{-\alpha}).$$

الشكل 4-4 يظهر الكثافة لتوزيع فريشيه لقيم مختلفة لـ α . الثلاث منحنيات على اليسار المبينة على الشكل تظهر دالة الكثافة عندما تكون $\alpha = 1$ و $\alpha = 2$ و $\alpha = 3$ (بقيم أعلى من α بوجود قمم أعلى).

كما أظهر الشكل 4-4 التوزيع للقيم القصوى لـ 16 رسم من التوزيع الذي له $F(x) = 1 - x^{-2}$ لـ $x > 1$. يجب أن يتسق هذا تقريباً مع توزيع فريشيه مع $\alpha = 2$ بقياس أفقي. بما أن $F(x) = 1 - kx^{-\alpha}$ مع $k = 1$ و $\alpha = 2$ فإن المقياس هو $a_N = (kN)^{1/\alpha} = 16^{1/2} = 4$. هذه المقارنة تم توضيحها أيضاً (المنحنيان بقمم تقترب من الـ 3): المقياس الملائم لدالة كثافة فريشيه تم توضيحها بخط منقط وتوزيع القيم القصوى بخط متصل. المقياس الأفقي لدالة التوزيع التراكمي سوف يتطلب مقياس رأسي متنسق لدالة الكثافة من خلال عامل للرقم 4 (حتى تظل دالة الكثافة تتكامل مع الـ 1). وهذا هو سبب انخفاض القمة هنا دون المقياس الطبيعي لفريشيه باستخدام $\alpha = 2$. الاتصال بين الخط المنقط والمتصل جيد جداً، خاصة بالنظر إلى الحجم الصغير لـ $N = 16$.

لدينا الآن صورة واضحة لما يحدث عندما يكون التوزيع الأصلي له مؤشر ذيل ثابت، ولكن ماذا سيحدث إن قمنا بتطبيق نفس النهج مع الذيل التي تتجه إلى الصفر بأسرع من متعدد الحدود (مثال، لهم نفس مؤشر الذيل غير المحدود)؟ أفضل طريقة لفهم هذا الوضع هو العمل من خلال مثال، ولهذا نضع في اعتبارنا التوزيع الأسّي، الذي لديه دالة توزيع تراكمي $F(x) = 1 - e^{-x}$ ودالة كثافة $f(x) = e^{-x}$ for $x > 0$ ويكون له نفس

خصائص مؤشر غير المحدود. مع وجود مؤشر ذيل محدود استخدمنا مقياس أفقي لـ $N = b(kN)^{1/\alpha}$ ولكن عندما تكون α لا نهائية يصبح هذا مقياس من خلال 1، مثال لا يوجد مقياس على الإطلاق. في حقيقة الأمر، سوف نقولك بتحويلها من خلال كمية معينة تعتمد على N بدلاً من المقياس الأفقي. في هذا المثال سنرغب في تحويلها بكمية $\log(N)$ ، لاحظ التالي:

$$F(\log(N) + x) = 1 - e^{-\log(N)} e^{-x} = 1 - \frac{\exp(-x)}{N}.$$

ولهذا

$$\begin{aligned} F_{\max}(\log(N) + x) &= (F(\log(N) + x))^N \\ &= \left(1 - \frac{\exp(-x)}{N}\right)^N. \end{aligned}$$

من خلال ملاحظتنا السابقة في المعادلة (4.2) حول الأسى كحد يمكننا رؤية أن N تصبح أكبر،

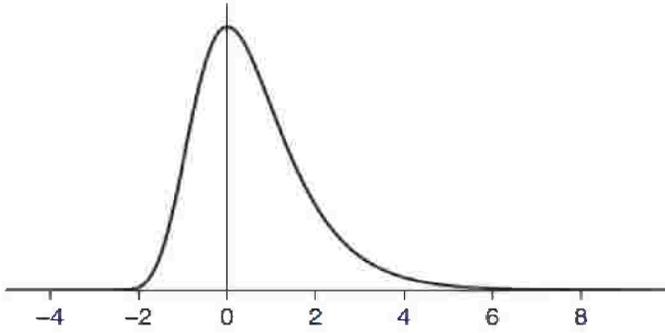
$$F_{\max}(\log(kN) + x) \rightarrow \exp(-e^{-x}).$$

من جديد يكون لدينا توزيع يصل إلى الحد. في حقيقة الأمر هذا التوزيع حينما يكون $F(x) = \exp(-e^{-x})$ يسمى بـ *Gumbel distribution* أي توزيع جامبل. وقد أظهرنا الحدود القصوى لـ 50 رسم من التوزيع الأسى حينما يكون $F(x) = 1 - e^{-x}$ يكون له توزيع يقترب من توزيع جامبل من خلال كمية تساوي $\log(50) = 3.91$.

قد يبدو من النظرة الأولى أن توزيع جامبل غير ممكن: له دالة توزيع تراكمي حينما يتم زيادة قيمة e إلى قيمة تكون هي نفسها قيمة أعلى.

دالة جامبل تكون

$$f(x) = (d/dx) \exp(-e^{-x}) = e^{-x} \exp(-e^{-x}).$$



الشكل 5-4: دالة الكثافة لتوزيع جامبل

يوضح الشكل 5-4 دالة الكثافة لجامبل ويمكننا رؤية أنه "عادي" جدًا (نظرا لجميع الأسس في تعريفها).

بالرغم من أن التوزيع الأسّي له مؤشر ذيل غير محدود فإنه لا يزال لديه منحنى أكثر سمكًا من التوزيع الطبيعي الذي يتضمن سلوك $\exp(-x^2)$ لدالة الكثافة. من الصعب التعامل مع التوزيع التراكمي للتوزيع، ولهذا سوف نستخدم توزيع أبسط ولكن مع استخدام خصائص $\exp(-x^2)$. ولهذا سنأخذ المثال الثاني للتوزيع مع دالة التوزيع التراكمي من خلال المحدود للقيم القصوى من الرسوم N سنحتاج إلى استخدام كلا المقياسين (من خلال العامل b_N) والتحويل (باستخدام الكمية a_N) وسنحدد

$$a_N = \sqrt{\log N} \text{ and } b_N = 1/(2\sqrt{\log N}).$$

يبدو أننا قد استخرجنا هذه القيم من b_N, a_N من اللاشيء، ولكننا سوف نوضح كيفية اختيارهما لاحقًا، والآن نلاحظ أن

$$\begin{aligned} F(a_N + xb_N) &= 1 - \exp(-(a_N + xb_N)^2) \\ &= 1 - \exp\left(-\left(\sqrt{\log N} + \frac{x}{2\sqrt{\log N}}\right)^2\right) \\ &= 1 - \exp\left(-\left(\log N + x + \frac{x^2}{4 \log N}\right)\right) \end{aligned}$$

$$= 1 - \frac{1}{N} \exp(-x) z_N(x),$$

حيث يكون

$$z_N(x) = \exp(-x^2/(4 \log N)).$$

وعليه

$$\begin{aligned} F_{\max}(a_N + b_N x) &= (F(a_N + x b_N))^N \\ &= \left(1 - \frac{z_N(x) \exp(-x)}{N}\right)^N. \end{aligned}$$

لاحظ أن القيم الثابتة لـ $z_N(x)$ ، x تقترب من الـ 1 حيث $N \rightarrow \infty$. ويعد هذا كافيًا لإيضاح

$$\lim_{N \rightarrow \infty} F_{\max}(a_N + x b_N) = \lim_{N \rightarrow \infty} \left(1 - \frac{\exp(-x)}{N}\right)^N = \exp(-e^{-x}). \quad 3-4$$

ونصل في نهاية المطاف إلى توزيع جامبل Gumbel: الذي هو بالضبط نفس توزيع الحد بالنسبة للحالة الأسية. قد يكون هذا مفاجئًا، لأننا لا نتوقع أن نصل إلى نفس التوزيع الحدية لسلكين ذيلين مختلفين: هناك اختلاف كبير بين $\exp(-x)$ و $\exp(-x^2)$ (على سبيل المثال، حينما تكون $x = 5$ سنحصل على $e^{-5} = 6.7 \times 10^{-3}$ و $e^{-25} = 1.4 \times 10^{-11}$). وجددير بالملاحظة أن المضاعف b_N يتجه إلى الصفر حيث $N \rightarrow \infty$. وهذا يتوافق مع مجموعة من التوزيع الذي لا يحدث مع التوزيع الأسي. حتى هنا هناك فرق نوعي بين الحالتين: مع ذيل أقل سمكًا المنوط بـ $\exp(-x^2)$ يمكننا أن نكون أكثر دقة في توقعنا للقيم القصوى لـ N من الرسوم من خلال شكل التوزيع.

المثالان الموضحين هما مثالان خاصان لنتيجة أكثر عمومية تسمى نظرية فيشر-تبيبت Fisher-Tippett theorem. هذه النظرية أظهرت لكافة التوزيعات ذو مؤشرات الذيل غير المحدودة (الأسفي في الأساس) أن توزيع جامبل يحدث كحد حين ننظر في الرسومات

المكررة. في حقيقة الأمر، الحالات التي يحدث هذا فيها هي حالات عامة يمكن افتراض حدوثها مع أي توزيع ذو مؤشر ذيل غير محدود. لكي نكون أكثر دقة نعرف a_N بأنها

$$1 - 1/N \text{ مجزأ للتوزيع (إحصائي)، على سبيل المثال:}$$

$$a_N = F^{-1}(1 - 1/N)$$

و

$$b_N = 1/(Nf(a_N)),$$

حيث f دالة الكثافة لـ F . ثم تظهر النظرية أنه عندما يكون مؤشر الذيل غير محدود (لا نهائي)،

$$\lim_{N \rightarrow \infty} F_{\max}(a_N + xb_N) = \exp(-e^{-x}).$$

سوف نعرض بيان رسمي لنظرية Fisher-Tippett بالمزيد من المناقشات في القسم الفرعي التالي.

مثال 4-4 تطبيق نظرية فيشر تبت Fisher-Tippett على التوزيع الطبيعي

سنطبق نظرية Fisher-Tippett الآن على دالة إيجاد التوزيع لأقصى 50 نموذج مستقل من توزيع طبيعي عادي. يكون له دالة كثافة

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

سنكتب $\Phi(x)$ لدالة التوزيع التراكمي العادية، $\int_{-\infty}^x \varphi(s) ds$. أول خطوة هي إيجاد القيم الصحيحة لـ a_N و b_N لـ $N = 50$ وسيكون لدينا

$$a_N = \Phi^{-1}(1 - 1/N) = \Phi^{-1}(0.98) = 2.054,$$

$$b_N = 1/(N\varphi(a_N)) = 1/(50\varphi(2.054)) = 0.413.$$

بهذه القيم يمكننا القول بأن $F_{\max}(a_N + b_N X)$ تقريباً هي توزيع جامبل، مثال

$$F_{\max}(2.054 + 0.413x) \approx \exp(-e^{-x}).$$

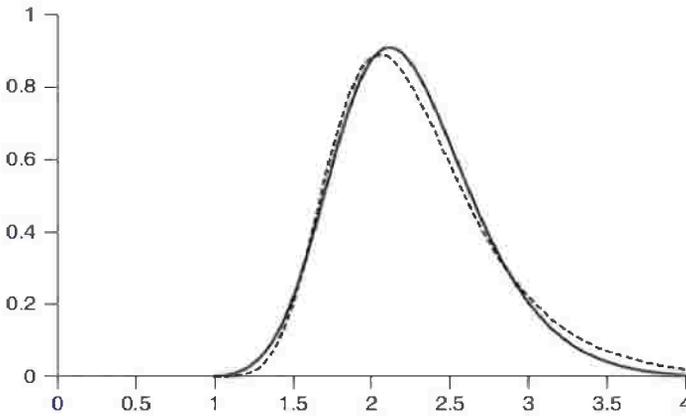
ويمكننا أيضًا عكس ذلك والقول بأن $F_{\max}(y)$ ، فرص أن تكون القيم القصوى لـ 50 نموذج أقل من y هي تقريبًا

$$F_{\max}(2.054 + 0.413x) \approx \exp(-e^{-x}).$$

يكون لها كثافة متوافقة

$$F_{\max}(2.054 + 0.413x) \approx \exp(-e^{-x}).$$

في الشكل 4-6 أظهرنا كثافة القيم القصوى لـ 50 رسم من توزيع طبيعي (الخط المتصل) مقارنة بتوزيع جامبل من خلال f (الخط المنقط). هذا لا يعد توافق جي، والذي يظهر أن التغير في هذه الحالة للتوزيع الطبيعي أكثر تباطؤًا. باستخدام مقياس ثابت a_N و b_N ويمكن أن يتحسن بشكل طفيف، ولكن لا يهم القيم المستخدمة، والتقريب لن يكون دقيقًا. (لاحظ أن القيم لـ a_N و b_N غير فريدة - الاختيارات المختلفة لهذا التسلسل يمكن أن يعطي نفس النتيجة في حد معين ويمكن أن يكون أفضل قليلاً للقيم الأقل لـ N .)



الشكل 6.4: مقارنة جامبل مع القيم القصوى لـ 50 نموذج من التوزيع الطبيعي القياسي

حتى هذه النقطة افترضنا أننا نتعامل مع توزيعات لها مجموعة لا نهائية. ما إذا كان مؤشر الذيل محدود أو لانهائي هو مسألة مدى سرعة كثافة الذي في الاتجاه إلى الصفر، ولكن افترضنا أن الكثافة لا تزال إيجابية. الآن ننتقل إلى حالة القيم القصوى للتوزيع. لذلك هناك نقطة نهاية ثابتة على الجانب الأيمن، x_{\max}

من الواضح أن التوزيع للقيم القصوى لـ N من النماذج من خلال هذا التوزيع سيكون له نفس هذا الوضع من القيم القصوى. الآن من المنطقية النظر في كيفية سرعة دالة التوزيع التراكمي للتوزيع الأصلي، $F(x)$ تقترب من الـ 1 حيث x تقترب من النقطة x_{\max} أو (المساوي) كيفية وتيرة اتجاه دالة الكثافة إلى الصفر. اتضح أن التوزيع للقيم القصوى يقترب أيضًا من نوع معين من التوزيع في هذه الحالة. في حقيقة الأمر، الحد يصبح توزيع وايبيل *Weibull distribution* والذي يكون له دالة توزيع تراكمي $H(x) = \exp(-(-x)^\alpha)$ for $x < 0$ لبعض الاختيارات لـ $\alpha > 0$. دالة الكثافة لـ Weibull هي $h(x) = \alpha(-x)^{\alpha-1} \exp(-(-x)^\alpha)$ for $x < 0$.

في حقيقة الأمر تم وصف هذه الحالة من خلال نظرية Fisher-Tippett. كما هو الحال مع الحالات الأخرى لدينا كمعاملات التحويل والتحجيم a_N و b_N حين تطبيقها. بما أن نظرية توزيع وايبيل Weibull تنتهي عند الصفر، التحويل يجب أن يكون بكمية x_{\max} . لذلك في نهاية المطاف يكون لدينا مجموعة من الأرقام b_N حينها يكون

$$\lim_{N \rightarrow \infty} F_{\max}(x_{\max} + xb_N) = \exp(-(-x)^\alpha) \text{ for } x < 0.$$

المقياس α لتوزيع وايبيل سيتم تحديده من خلال سلوك F قرب x_{\max} . حال، على سبيل المثال، $F(x) = 1 - k(x_{\max} - x)^2$ لبعض K ثم التوزيع المحدود لوايبيل مع $\alpha = 2$.

1-2-4* المزيد من التفاصيل على التوزيعات القصوى ونظرية Fisher-Tippett

في هذا القسم سنوضح بعض التفاصيل الرياضية المفقودة، نبدأ بالسؤال الدائر حول أسباب أن يكون لدى التوزيع مؤشر ذيل له قيم قصوى (من N نماذج مستقلة) تقترب من توزيع فريشيه.

النقاش الذي وضعناه في القسم الفرعي السابق يفترض شكل معين لدالة التوزيع التراكمي ولكن تعريف مؤشر الذيل أكثر عموماً من هذا.

نحتاج إلى العمل على دالة متغيرة بوتيرة بطيئة. افترض أن $L(t) = (1 - F(t))t^\alpha$ دالة ذو متغير بطيء لـ t (هذا يعني أن α مؤشر ذيل لـ F ولكن إدعاء أقوى قليلاً). ولهذا، $\lim_{z \rightarrow \infty} L(tz)/L(z) = 1$ لأي قيمة لـ $t > 0$. في هذا التعبير سنحدد $z = b_N$ و $t = x$ وسنرى حين تعريفنا

$$z_N(x) = \frac{(1 - F(b_N x))b_N^\alpha x^\alpha}{(1 - F(b_N))b_N^\alpha} = \frac{(1 - F(b_N x))x^\alpha}{(1 - F(b_N))}, \quad (4.4)$$

ثم الحالة التي سيكون فيها هناك متغير بطيء يعني أن أي قيمة ثابتة $1 \rightarrow z_N(x)$ توفر $b_N \rightarrow \infty$ الآن سنحدد

$$b_N = F^{-1}\left(1 - \frac{1}{N}\right),$$

حينها يكون $1 - (1/N)$ تجزيء للتوزيع، والاختيار الصحيح لعامل القياس. هذا سوف يقترب بالفعل ∞ حيث تصبح N أكبر. والآن نلاحظ أننا يمكن أن نعيد ترتيب المعادلة (4.4) للحصول على

$$\begin{aligned} 1 - F(b_N x) &= z_N(x)(1 - F(b_N))x^{-\alpha} \\ &= \frac{z_N(x)x^{-\alpha}}{N}. \end{aligned}$$

ولهذا

بسبب سلوك الحد لـ Z_n نعرف أن أي $\varepsilon > 0$ سيكون لها $1 - \varepsilon < z_N < 1 + \varepsilon$ لكون N كبيرة كفاية. ولهذا

$$\lim_{N \rightarrow \infty} \left(1 - \frac{(1 - \varepsilon)x^{-\alpha}}{N}\right)^N > \lim_{N \rightarrow \infty} F_{\max}(b_N x) > \lim_{N \rightarrow \infty} \left(1 - \frac{(1 + \varepsilon)x^{-\alpha}}{N}\right)^N.$$

بما أن الجانب الأيمن والأيسر يمكن أن يتم تقريبيهما كما نرغب لـ $\exp(-x^{-\alpha})$ من خلال اختيار قيمة صغيرة كفاية لـ ε يمكننا خلق

$$\lim_{N \rightarrow \infty} F_{\max}(b_N x) = \exp(-x^{-\alpha}),$$

وهذا هو توزيع فريشيه الذي كنا نعمل عليه.

هذه المناقشات التي تضيق حدود F_{\max} بين التعبيرين اللذان يمكن استخدامها بالقرب الذي نرغب به للدالة الأسية هي أيضًا ما نحتاج إلى إظهاره وهو أن الحد في المعادلة صحيح في الشكل (3-4).

الآن نريد أن نتقل إلى وصف رسمي أكثر لنظرية **Fisher-Tippett**. لاحظ أنه عندما نقول أن توزيع القيم القصوى يقترب من توزيع محدود، يكون ما نعنيه هو التوزيع الحاصل من خلال عملية وضع قياس وتحويل (بهذا القياس والتحويل المعتمد على عدد N للرسوم من التوزيع الأصلي). ولهذا أي شيء تم وضع قياس له وتم نقله من توزيع محدود يمكن استخدامه. ونستخدم مصطلحات مثل: توزيعين F_1 و F_2 وهما نفس النوع في حال كان من الممكن الحصول عليهما من إحداهما من خلال القياس والنقل. مثال، إذا كان

$F_1(x) = F_2(a + bx)$ للاختيار الصحيح لـ a و b . على سبيل المثال، أي توزيع طبيعي بغض النظر عن متوسطه أو انحرافه المعياري هو من نفس النوع.

نقول أن التوزيع F في أقصى مجال لجذب التوزيع H في حال كان هناك تسلسل لـ a_N و b_N مع $N \rightarrow \infty$

$H(x)$ لكل x . نكتب ذلك كالتالي $F \in MDA(H)$. المصطلح هنا يمكن أن يكون مضلل - المجال الأقصى للجذب يبدو كأنه أعلى مجال للجذب بطريقة ما، ولكن بدلاً من ذلك فإن المعنى الحقيقي لذلك هو مجال الجذب تحت عوامل قصوى. القول بأن $F \in MDA(H)$ هي نفسها بالضبط حين القول الحد الأقصى لـ N للنماذج من F لها توزيع يكون له نفس الشكل H (بعد تحجيمها بـ b_N ونقلها بـ a_N).

ونظرًا لأنه تم اختيار a_n و b_n بشكل عشوائي، فإن هذا بيان حقيقي حول أنواع التوزيعات: في حال كانت F توزيع لنفس النوع ك F و H من نفس النوع ك H ثم $F \in MDA(H)$ و $F \sim F$ مساويين بالضبط للبيان.

الآن نحن على استعداد لإقرار نظرية Fisher–Tippett في الصيغة المعتادة لها.

النظرية 1-4 في حال كان F في $MDA(H)$ لبعض التوزيعات H (و H غير مركزة في نقطة بعينها) إذا هناك مقياس ξ حينما تكون H من نفس النوع ك

$$H_{\xi}(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}) & \text{for } \xi \neq 0, \\ \exp(-e^{-x}) & \text{for } \xi = 0. \end{cases}$$

الصيغة لـ H_{ξ} تُعرف بأنها مجموعة من التوزيعات التي عادة ما تسمى بالقيم القصوى العامة *generalized extreme value* أو توزيعات GEV. وتنطبق هذه الصيغة أيضًا على النطاق الذي يحمل قيم x ، وبما أننا نحتاج إلى $H_{\xi}(x)$ صعودًا من 0 إلى 1. عندما $\xi < 0$ نحتاج $\xi \leq -1$ وعندما $\xi > 0$ نحتاج $\xi \geq 1$.

هناك ثلاث حالات للنظر وفقًا لما إذا كانت قيمة ξ إيجابية أو صفر أو سلبية. الحالة التي تكون فيها $\xi = 0$ هي نظرية توزيع جامبل. حينما تكون $\xi > 0$ نستطيع أن نضع $\alpha = 1/\xi$ ونحصل على دالة توزيع $H(x) = \exp(-(1 + x/\alpha)^{-\alpha})$ ، والتي هي من نفس نوع نظرية توزيع فريشيه $\exp(-x-\alpha)$ (نحن فقط بالحاجة إلى تحجيمها بمقدار $1/\alpha$ وتحريكها بمقدار 1). السبب في الحصول على هذه النتيجة في هيئة أكثر تعقيدًا بـ ξ بدلًا من α هي أنها تجعلها أكثر وضوحًا بالحالة التي تكون فيها $\xi = 0$ التي يمكن الوصول إليها في الحدود $0 \rightarrow \xi$ سواء من الأعلى أو الأسفل. في كلا الحالتين نستخدم فقط ملاحظتنا العادية، المعادلة (4.2) وهي أن $(1 + x/n)^n$ تتسق مع e^x as $n \rightarrow \infty$.

نحن بالحاجة مناقشة هذا الأمر بمزيد من التفصيل حينما يكون $\xi < 0$. في هذه الحالة توزيع GEV أو توزيع القيم القصوى يكون له نقطة نهاية ثابتة على الجانب الأيمن، وبما أن $\exp(-(1 + \xi x)^{-1/\xi})$ تصل إلى القيمة القصوى لـ 1 حينما تكون $\xi = -1$ ، وتحديد الجانب

الأيمن للتوزيع. من غير الصعب رؤية أن توزيع القيم القصوى العامة GEV من نفس نوع توزيع وايبل. لخلق صلة بين المقياس α لسلوك دالة التوزيع الأصلي، يجب علينا العودة إلى دالة متغير بطيء. بالرغم من ذلك، فكرة دالة المتغير البطيء تتمحور حول كيفية سلوك الدالة باتجاهها إلى ما لا نهاية، ولهذا يجب علينا القيام ببعض التلاعب والعمل على المعكوس للاختلافات بين X و x_{max} في حال

$$(1 - F(x_{max} - (1/z)))z^\alpha$$

كانت دالة متغير بطيء ثم القيم القصوى لـ N من الرسوم من التوزيع تقريباً (بعض التحجيم والتحرك) نظرية وايبل مع مقياس α (او GEV مع مقياس $-1/\alpha$) نظرية Fisher-Tippett في أصلها لا توضح مدى احتمالية وجود توزيع محدود أو تحجيمي لـ F_{max} ولكنها تحدد فقط شكل هذا التوزيع في حال حدوثها. ولكن في حقيقة الأمر، يعد إيجاد أي مثال دون توزيع محدود أو تحجيمي صعباً للغاية.

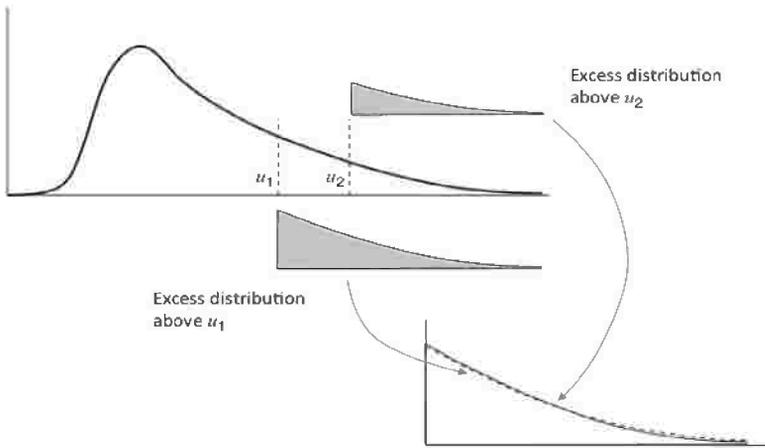
تذكر أننا نسمح بالتحجيم والتحرك ولهذا يعد شكل التوزيع هو المهم في النهاية. يمكننا على سبيل المثال، استخدام التحجيم والتحرك لإصلاح النقط المئوية الخامسة والمئوية الخامسة والتسعين. ثم مشاهدة ما سوف يحدث إلى نقاط التجزئة الأخرى مع تزايد N . وفي حال تغاير جميعهم سيكون لدينا توزيع محود (تحجيمي).

إذا تخيلنا أنه لا يوجد توزيع تحجيمي أو محدود سيكون من الصعب رؤية قدرتنا على التنقل بين أشكال مختلفة، ولكن إذا لم يكن هناك تنقل إذا النقاط المجزئة المحصورة بين النسب المئوية الخامسة والخامسة والتسعون الملزمة للتغاير. ولهذا فإن الأمر متعلق بالقيم القصوى التي يمكن أن تعطي نتائج خاطئة، مثال من خلال أن يكون لدينا منحني x^{-k} مع تزايد قيمة k أكثر وأكثر مع تزايد N ولكن فعل ذلك بطريقة لا تؤدي إلى نوع من أنواع التوزيع المحدود. اتضح أن ذلك لن يحدث للتوزيع العملي، ويمكن إيضاح ذلك من خلال النظر بشكل أكثر تفصيلاً في الحالات الضامنة أن F في $MDA(H_\xi)$. قمنا بالفعل إظهار كيفية حالة من المتغير البطيء كافي حينها يكون هناك مؤشر ذيل لـ α (ونفس الحالة تنطبق على حد

ثابت على الجانب الأيمن) ولكننا بحاجة إلى شيء أكثر تعقيداً للتعامل مع حالة يكون فيها $\xi = 0$ ولن نقوم بعرض أية تفاصيل هنا.

3-4 توزيعات زائدة

الآن ننتقل إلى موضوع مختلف ذو صلة، فبدلاً من النظر للقيم القصوى لـ N من المتغيرات العشوائية، في هذا القسم سوف ننظر بشكل مباشر في سلوك منحني التوزيع. الفكرة مشابهة للمناقشات الواردة في القسم السابق، حينما انصبي اهتمامنا على شكل التوزيع للقيم القصوى. ما الآن ينصب اهتمامنا على شكل التوزيع فوق الهدف u . وكلما زاد هذا الهدف نتساءل ما إذا كان ذلك توزيع محدود للزائد (مثال، جزء من التوزيع أعلى الهدف u) الذي يفترض دائماً بأننا نحجم بشكل ملائم. تم شرح الفكرة في الشكل 7-4 حيثما كان هدفين مختلف للقيم u لهم توزيعات زائدة تشبه بعضها بعضاً.



الشكل 7-4: التوزيعات أعلى الهدف u_1 والتوزيعات أعلى الهدف u_2 يمكن أن يتم تحجيمها بشكل مشابه

التوزيع الزائد تم الحصول عليه من خلال التوزيع $X - u$ لبعض أهداف u ، المشروط بـ X أكبر من u . نكتب التالي كـ $Fu(x)$ وعليه

$$F_u(x) = \Pr(X - u \leq x | X > u), \text{ for } x > 0.$$

هذه هي ببساطة احتمالية أن يكون المتغير العشوائي X في نطاق u إلى $x + u$ من خلال

كونها أكبر من u مثال قاعدة بايز Bayes

$$F_u(x) = \frac{F(x + u) - F(u)}{1 - F(u)}, \text{ for } x > 0.$$

هذا التعبير لـ F_u يمكن كتابته كالتالي

$$F_u(x) = 1 - \frac{1 - F(x + u)}{1 - F(u)}.$$

سوف نستعرض ذلك العديد من التوزيعات F وسوف يكون هناك توزيعات محدودة (معرفة في التحجيم) لـ F_u تصبح كبيرة جداً. هذه النظرية تتسق بشكل كبير مع نظرية التوزيع للقيم القصوى لـ N من النماذج الموضحة في القسم السابق.

لشرح ذلك الأمر، افترض أن $F(x) = 1 - kx^{-\alpha}$ ، حتى F يكون لها مؤشر ذيل لـ α .

ثم

$$\begin{aligned} F_u(x) &= 1 - \frac{k(x + u)^{-\alpha}}{k(u)^{-\alpha}} \\ &= 1 - \left(1 + \frac{x}{u}\right)^{-\alpha}, \text{ for } x > 0. \end{aligned}$$

وهكذا، إذا استخدمنا مقياس u نجد أن:

$$F_u(ux) = 1 - (1 + x)^{-\alpha}$$

حيث $x > 0$

وهنا نجد أنه ليس هناك تقريب، فإن التوزيع الزائد (بعد القياس) يعطى بهذه الصيغة بالضبط. ويسمى هذا التوزيع بتوزيع باريتو النوع الثاني. ويوضح القسم الثاني أن هذا الشكل من التوزيع المحدود هو الوحيد الذي يمكن أن ينشأ من خلال مؤشر المنحنى a

ويمكننا تنفيذ نفس النوع من الحساب عندما تكون توزيعات أرباح لها مؤشر منحني لانهائي. لذلك نحن نفترض ذلك:

$$F(x) = 1 - \exp(-x^2).$$

فإن

$$\begin{aligned} F_u(x) &= 1 - \frac{\exp(-(x+u)^2)}{\exp(-u^2)} \\ &= 1 - \frac{\exp(-x^2 - 2ux) \exp(-u^2)}{\exp(-u^2)} \\ &= 1 - \exp(-x^2 - 2ux), \text{ for } x > 0. \end{aligned}$$

وبالنظر في التحجيم بنسبة $1 / (2u)$ ، نجد أن:

$$F_u\left(\frac{x}{2u}\right) = 1 - \exp\left(-\frac{x^2}{4u^2}\right) \exp(-x), \text{ for } x > 0.$$

في حين تصل $\exp(-x^2/(4u^2))$ إلى 1، حيث $u \rightarrow \infty$ ، نجد أن:

$$F_u\left(\frac{x}{2u}\right) \rightarrow 1 - \exp(-x), \text{ for } x > 0.$$

حيث $u \rightarrow \infty$ ، وبالتالي في هذه الحالة فإن استخدام التحجيم بنسبة $1 / (2u)$ يعطي التوزيع الأسّي بشكل تقريبي. واتضح أن الاختيار الصحيح للتحجيم المستمر سيحقق توزيع أسّي في الحد الأقصى لأي توزيع معقول له مؤشر منحني لانهائي. ويظهر توزيع باريتو في الحد الأقصى عندما يكون مؤشر المنحني محدود ويحدث التوزيع الأسّي عندما يكون مؤشر المنحني لانهائي. وهذا يذكرنا بنظرية فيشر - تيببت لتوزيع الحد الأقصى لعينات N من نفس التوزيع. في الواقع، وكما ناقش في القسم التالي، فإن شروط التقارب هي نفسها تماما.

مرة أخرى يمكننا تلخيص الموقف النهائي في نظرية واحدة، ولكن هذه المرة نستخدم توزيع باريتو المعمم لالتقاط أشكال التوزيع المختلفة التي قد تحدث. وستحدث عن هذه

النظرية بمزيد التفاصيل في مناقشة القسم الفرعي التالي. ولكن في الوقت الراهن نريد فقط أن نفهم كيفية استخدام النتيجة. وسنعمد على القيمة الكبيرة لـ u والنسبة التقريبية لـ $F_u(x)$ وتوزيع باريتو المعمم الذي له وظيفة التوزيع التالية:

$$G_{\xi, \beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{for } \xi \neq 0 \\ 1 - \exp(-x/\beta) & \text{for } \xi = 0 \end{cases} \quad (5-4)$$

حيث أن $\beta < 0$ و $\xi \geq 0$ ، ومن ثم يجب أن تكون $x \geq 0$ ، حيث تتوافق الحالة $\xi < 0$ مع توزيع F والقيمة القصوى x_{\max} . وهنا يتطلب أن تكون $x \leq -\beta/\xi \geq 0$.

وبشكل عام فإن المعلمة ξ تلتقط الشكل وتعد β بمثابة التحجيم الثابت. وتكون المعلمة ξ هي عكس مؤشر المنحنى، وبالتالي فإن $\xi = 1/\alpha$. وسبب استخدام ξ بدلا من α هو لأن صيغة باريتو المعممة تتسق مع قيم ξ المختلفة. حيث يتم إيجاد $\xi = 0$ في الحد الذي تصل ξ إلى الصفر.

إذا كانت F تحتوي على مؤشر منحنى $\alpha < \infty$ (أي $\xi > 0$) فيكون لدينا:

$$F_u(x) = 1 - \left(1 + \frac{x}{u}\right)^{-\alpha}, \text{ for } x > 0.$$

يمكننا مقارنة هذا التعبير مع توزيع باريتو المعمم الموضح في المعادلة (5-4). ويظهر التحويل من ξ إلى α أن معلمة التحجيم β تعطى بواسطة $u\xi$ في هذه الحالة. وفي حالة وجود a لانهائية (أي $\xi = 0$) سيتم تحديد معلمة التحجيم بواسطة شكل المنحنى. وهو المثال الذي تمت مناقشته أعلاه $f(x) = 1 - \exp(-x^2)$ حيث قيمة β هي $1/(2u)$. وسوف ستعطي أمثلة أخرى لتوزيعات مؤشر المنحنى اللانهائي قيم مختلفة لـ β . وتعطى دالة الكثافة لتوزيع باريتو المعمم من خلال:

$$g_{\xi, \beta}(x) = (1/\beta)(1 + \xi x/\beta)^{-(1/\xi)-1}$$

فمن السهل رؤية التوزيع في المعادلة (5-4). وتعطى الكثافة عند $\xi = 0$ في المعادلة

$$g_{\beta}(x) = (1/\beta) \exp(-x/\beta)$$

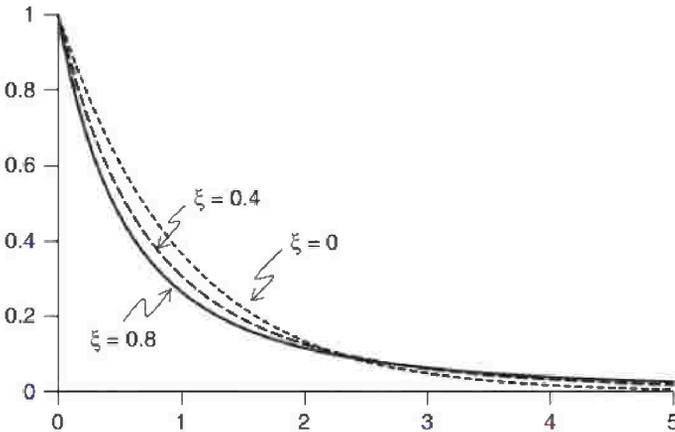
والذي يعد الحد لـ $g_{\xi,\beta}(x)$ حيث تصل ξ إلى الصفر. ويحدد الشكل 4.8 هذه الكثافة لقيم مختلفة لـ ξ ، في كل حالة تقدر فيها قيمة β بـ 1.

إذا كان الزائد لديه توزيع باريتو المعمم، فمن الطبيعي أن نسأل عن متوسط التوزيع: سيكون هذا هو متوسط قيمة الفائض على u ، أي $E(X - u | X > u)$. وعندما تكون X موزعة مثل $G_{\xi,\beta}$ يتضح أن قيمة المتوسط تعطى بواسطة

$$E(X) = \beta / (1 - \xi).$$

يطلب التمرين 4-4 أن تقوم بتنفيذ التكامل لإظهار أن هذه الصيغة صحيحة. ويتم استخدام الصيغة نفسها عندما تكون $\xi=0$ حيث أن التوزيع الأسي بكثافة $(1/\beta) \exp(-x/\beta)$ له متوسط β . لاحظ أن الاختلافات في الشكل 4-8 بين التوزيعات الثلاثة لا تبدو هائلة، ولكن يظهر متوسط الكثافة على الخط متين وهو 5 ومتوسط الخط المتقطع هو $3/5$ ومتوسط الخط المنقط هو 1.

في حال كان $\xi \geq 1$ إذن الصيغة لم تعد منطقية (إعطاء نتيجة سلبية أو لانهائية). هذا هو بالضبط ما نتوقعه: مؤشر الذيل هو $\xi / 1$ ، وإذا كان هذا المؤشر أقل من 1 يكون المتوسط غير موجود.



الشكل 8-4: دالة الكثافة لتوزيع باريتو المعمم مع بارامترات مختلفة. جميع قيم β هي 1

بالنظر إلى التوزيع المعروف من خلال F يوصف توزيع باريتو التوزيع الفائض فوق الحدود u . الآن نستكشف ماذا يحدث عندما يتم استخدام هدف آخر. افترض أن F_u معطاه. إذا الصيغة التي تعرف F_u هي

$$1 - F_u(x) = \frac{1 - F(x + u)}{1 - F(u)}$$

ويمكن إعادة كتابة ذلك كالتالي

$$1 - F(x + u) = (1 - F_u(x))(1 - F(u)). \quad (6-4)$$

افترض أن لدينا $a > u$ ، إذاً، من هذه المعادلة (6.4) يمكن استنتاج التالي

$$\begin{aligned} 1 - F_v(x) &= \frac{1 - F(x + v)}{1 - F(v)} \\ &= \frac{(1 - F_u(x + v - u))(1 - F(u))}{(1 - F_u(v - u))(1 - F(u))} \\ &= \frac{(1 - F_u(x + v - u))}{(1 - F_u(v - u))}. \end{aligned}$$

في الحالة التي يكون فيها $\xi \neq 0$ و F_u معطى من خلال توزيع باريتو المعمم، نحصل على

$$1 - F_v(x) = \frac{(1 + \xi(x + v - u)/\beta)^{-1/\xi}}{(1 + \xi(v - u)/\beta)^{-1/\xi}} = \left(1 + \frac{\xi x}{\beta + \xi(v - u)}\right)^{-1/\xi}.$$

تعد هذه صيغة أيضاً لتوزيع باريتو المعمم ولكن مع بارامتر β محل محله $\xi(v - u) + \beta$. بمعنى آخر، زيادة الحدود بكمية $\Delta = v - u$ يترك شكل البارامتر ξ دون تغيير، ولكنه يزيد من مقياس البارامتر بكمية $\Delta \xi$.

في الحالة التي يكون فيها $\xi = 0$ و F_u معطاه من خلال صيغة توزيع باريتو المعمم، يكون لدينا:

$$1 - F_v(x) = \frac{\exp((x + v - u)/\beta)}{\exp((v - u)/\beta)} = \exp(x/\beta).$$

حينما يكون $\xi = 0$ ، فإن زيادة الحدود لا يشكل فارقاً للتوزيع الفائض، ولكنه يستقر كأسي مع نفس الباراميتير. هذه هي خاصية الذاكرة المعروفة من التوزيع الأسي: إذا كان الوقت بين الأحداث هو توزيع أسي إذاً معرفة أن شيئاً لم يحدث حتى الآن لا يشكل فارق في توزيع الوقت إلى الحدث التالي.

هذه النتيجة حول الطريقة التي تغيرت بها β (أو لم تتغير) مع وجود حاجة إلى معاملة الحدود بعناية، لأنه يفترض أن التوزيع الزائد هو بالضبط توزيع باراميتو المعمم وهو مجرد تقريب. على سبيل المثال، في مناقشتنا في وقت سابق حول تلك المسألة التي يكون فيها $F_u(x/2u) = 1 - \exp(-x^2)$ سنجد أن أفضل التوزيع الفائض F_u له خاصية أن $F_u(x/2u)$ يقرب من دالة التوزيع التراكمي للتوزيع الأسي. بمعنى آخر، $F_u(x)$ معطى من خلال تقريباً $G_{\xi, \beta}(x)$ مع $\beta(u) = 1/(2u)$ و $\xi = 0$. ولهذا في هذه الحالة نلاحظ وجود تغيير في β حيث تتغير u حتى وإن كان من المفترض أن تكون β ثابتة في هذه الحالة. بالرغم من ذلك، الموضح كمضاعف للتغير في u يكون الاختلاف ضئيل. إذا أخذنا $v > u$ وتركنا $\Delta = v - u$ يكون لدينا

$$\beta(v) = \frac{1}{2v} = \frac{1}{2u} + \frac{u - v}{2uv} = \beta(u) - \frac{1}{2uv} \Delta.$$

1-3-4 مزيد من التفاصيل حول تجاوزات الحدود

أولاً نعود إلى السؤال حول سلوك التوزيع الفائض حينما يكون لدى F مؤشر ذيل a . كما فعلنا سابقاً، نستخدم فكرة أقوى لدالة متغيرة بطيئة ونفترض أن $(1 - F(t))t^\alpha$ دالة متغير بطيء لـ t (يعني هذا أن a مؤشر ذيل لـ F). ولهذا في حال

$$z_u(k) = \frac{(1 - F(ku))u^\alpha k^\alpha}{(1 - F(u))u^\alpha} = \frac{(1 - F(ku))k^\alpha}{(1 - F(u))},$$

إذاً، لأي ثابت من $k, z_u(k) \rightarrow 1$ as $u \rightarrow \infty$ ولهذا

$$\begin{aligned} F_u(ux) &= 1 - \frac{1 - F(ux + u)}{1 - F(u)} \\ &= 1 - (1 + x)^{-\alpha} z_u(1 + x), \end{aligned}$$

حينما يكون لدينا التعريف $Z_u(K) = 1 + X$. ولهذا حيث $u \rightarrow \infty$ يكون لدينا

$$F_u(ux) \rightarrow 1 - (1 + x)^{-\alpha} \text{ for } x > 0. \quad (7-4)$$

يمكننا القول بشكل عام أن X يتم توزيع حسب توزيع باريتو $Pa(\alpha, \kappa)$ في حال الحصول على دالة التوزيع التراكمي من خلال

$$F(x) = 1 - \left(1 + \frac{x}{\kappa}\right)^{-\alpha},$$

ولهذا، في هذه الحالة أظهرنا أنه بعد القياس من خلال u ، التوزيع الفائض هو باريتو Pareto مع $k=1$.

الآن سنعطي المساوي لها لنظرية Fisher–Tippett في هذه الإعدادات والتي تظهر دول توزيع باريتو المعمم الموضح أعلاه.

نظرية 2-4: في حال كان F في $MDA(H_\xi)$ إذا يكون هناك دالة لـ $\beta(u)$ مع

$$\lim_{u \rightarrow \infty} F_u(\beta(u)x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - e^{-x} & \text{for } \xi = 0. \end{cases}$$

لاحظ أنه لا يوجد تشابه بين الصيغتين بين تلك النتائج ونظرية Fisher–Tippett فحسب بل أنها نفس مجموعة التوزيعات التي تتقارب مع الحد الأقصى لتوزيع الصيغة H_ξ التي يتضح أن لها توزيع فائض يتقارب مع $G_{\xi, \beta}(x)$ لبعض خيار الدالة $\beta(u)$.

مع هذه الصيغة نحصل على التوزيع الأسّي في الحد كما $\xi \rightarrow 0$ ، وبالتالي فإن شكل التوزيع يتغير بطريقة مستمرة مع ξ . ومع ذلك، لأن هناك معامل ترجيحي $\beta(u)$ ، يمكننا وضع قياس x بعامل آخر من ξ ونعود إلى النموذج الذي قدمناه في وقت سابق، بمعنى 1

$(1+x)^{-\alpha}$ - كحد أقصى. وبالنسبة إلى $\xi \geq 0$ ، يحدد التوزيع المقيد على المدى $x \geq 0$ وعندما $0 \leq x \leq -1/\xi$ نحتاج إلى $\xi < 0$

ومن المثير للاهتمام مقارنة هذه النتيجة مع نظرية فيشر-تبيبت Fisher-Tippett. هناك نعمل مع أقصى قدر من N من العينات، وننظر إلى الحد كما $N \rightarrow \infty$ في حين هنا نعتبر وجود زيادة على u وننظر إلى الحد كما $u \rightarrow \infty$. في نظرية فيشر-تبيبت، فإن دالة التوزيع التراكمي المحدودة لها الشكل $\exp(-Z(x))$ حيث تتحرك دالة التوزيع التراكمي من 0 إلى 1 حيث ينتقل التعبير $Z(x)$ من ∞ إلى 0 ، في حين أن التوزيع المحدود هنا هو ببساطة $1 - Z(x)$ وتعرف على المدى الذي يتحرك فيه Z من 1 إلى 0 . وعلاوة على ذلك، في نظرية فيشر-تبيبت قد تحتاج إلى كل من التحجيم والتحويل (أي كل a_N و b_N وغير الصفرية) في حين أننا هنا فقط بحاجة إلى مقياس بواسطة $\beta(u)$.

4-4 التقدير باستخدام نظرية القيمة القصوى

إن فهم ذيل التوزيع له قيمة خاصة إذا أردنا تقدير الاحتمالات المرتبطة بالأحداث التي لم نشاهدها، أو نادرا ما نراها. والفكرة هنا هي تقدير شكل وحجم المعلمات أو البارامترات من البيانات ومن ثم استخدامها لتقدير احتمال الفائدة. بطبيعة الحال، سوف نحتاج إلى افتراض أن هناك اتساقا كافيا في سلوك الذيل الذي تنطبق عليه النظرية التي ناقشناها، ولكن هذا هو التحذير الذي يجب أن يوضع في الاعتبار مهما كان التقدير الذي نقوم به. التقدير قدي يعني استنتاج ما قد يحدث في مرحلة تخلو من البيانات، وهذا الاستنتاج لا بد أن نستفيد منه من بعض افتراضات النمذجة.

يعيدنا ذلك إلى السيناريو الافتتاحي لهذا الفصل، الذي يتمثل فيه الشرط في تقدير كمية من توزيع الاحتياجات الأسبوعية لدواء معين. في هذا المثال، أرادت ماريا أن تعرف قيمة x التي سيتم تجاوزها في أسبوع واحد فقط في 156 أسبوع. وهو تقدير يتجاوز الـ 104 نقطة من البيانات بكثير التي لديها، وبالتالي فإن أي تقدير سيكون غير مؤكد. ومع ذلك فإن الأفكار التي نستكشفها هنا تسمح لنا باستخدام افتراضات غير مقيدة تماما لنقول شيئا ما يحدث بشكل جيد في الذيل. نحن لسنا بحاجة إلى استخدام نموذج معين

للتوزيع (سواء كان ذلك طبيعيا أو أسيا أو أي شيء آخر) لأن التقديرات التي نجريها سوف تنطبق على كل هذه التوزيعات وغيرها. كل ما علينا أن نفترض هو أن هناك ما يكفي من الانتظام في السلوك في الذيل - وجود مؤشر ذيل معين سيكون كافيا.

إن فكرة استخدام افتراضات على سلوك منتظم بشكل معقول من أجل إجراء استقراء ليست مختلفة جدا من حيث نوع ما نقوم به عند استيفاء البيانات (أي عندما نقدر السلوك بين نقاط البيانات الموجودة). التقدير أو الاستقراء يفترض أن سلوك التوزيع لا يتغير فجأة للقيم المتطرفة، في حين جعل التقديرات من ضمن مجموعة من القيم التي لاحظناها يجعل افتراض أنه لا يوجد تغيير مفاجئ في السلوك داخل هذا النطاق. الفرق هو أنه عندما نقوم بالاستقراء بحكم التعريف، لا شيء يوجد في البيانات يُمكن أن يحذرنا عن التغييرات، في حين قد يتم الكشف عن تغيير كبير في التوزيع عند نقطة ضمن مجموعة من البيانات نلاحظها من خلال النظر في البيانات. وقد يبدو ذلك تمييزا مهما، ولكننا نحتاج في الواقع إلى الكثير من البيانات لنكتشف مثل هذا التغيير.

المشكلة التي تواجه ماريا، هي أن المشكلة ستظل قائمة حتى مع وجود أربع أو خمس سنوات من البيانات. إذا كان هناك لسبب ما، تغيير في سلوك توزيع المتطلبات الأسبوعية التي تحدث عند نقطة 99٪، ثم النظر في البيانات التجريبية حيث من المتوقع أن تكون أكبر من اثنين أو ثلاث نقاط من البيانات من هذه القيمة لن تكشف أبدا ما يجري. لذلك، فإن مضاعفة كمية البيانات سيساعد بالتأكيد في إجراء تقدير جيد، ولكن سيكون هناك حاجة إلى افتراض بعض الانتظام في البيانات.

في هذا القسم سوف نقدم طريقة من ثلاث خطوات لتقدير القيمة المعرضة للخطر أو العجز المتوقع (أو قياس المخاطر الأخرى). ومع ذلك، فمن المهم أن ندرك أن هذه مجرد واحدة من عدة طرق التي يمكن استخدامها. ولا سيما فيما يتعلق ببيانات السلاسل الزمنية، هناك حجج جيدة لاستخدام نهج أكثر تعقيدا يأخذ في الاعتبار التغيرات في التقلبات على مر الزمن (كما هو الحال في كثير من الأحيان مع البيانات المالية). ومع ذلك، عندما لا تكون الارتباطات بين نقاط البيانات قوية جدا، وعندما لا يكون هناك سبب لتوقع حدوث

تغييرات في التوزيع الأساسي مع مرور الوقت، فإن الطريقة التي سنصفها يمكن أن تكون فعالة جدا.

1-4-4 الخطوة الأولى اختيار حد لـ α

الفكرة هي تقريب توزيع الخسائر فوق مستوى هدف معين α ، وذلك باستخدام توزيع باريتو المعمم. نحن بحاجة لجعل قيمة α كبيرة بشكل معقول بحيث يتم تطبيق نظرية القيمة القصوى. ومع ذلك، بما أننا سننتهي في نهاية المطاف بتقدير معلمات توزيع باريتو المعمم من البيانات التي تحدث فوق الحدود، يجب علينا التأكد من وجود نقاط بيانات كافية للقيام بذلك. ليس من الصعب جدا اختيار قيمة معقولة للهدف مع مجموعة كبيرة من البيانات، ولكن يمكن أن يكون صعباً إذا كانت مجموعة البيانات صغيرة.

النهج الذي سنوضحه يستخدم عينة مخطط لمتوسط بسيط. وبالنسبة إلى أي قيمة للهدف v ، فإننا ننظر إلى المتوسط الذي يتجاوز فيها نقاط البيانات هذا الحدود، حيث يبلغ متوسطه أعلى بقليل من نقاط البيانات التي هي أكبر من v . متوسط الزيادة في العينة هو تقدير للقيمة الحقيقية، وكما رأينا بالفعل، يكون هذا $(1 - \beta)/\beta$ إذا كان التوزيع الزائد لتوزيع باريتو المعمم $G_{\xi, \beta}$ وتبين المناقشة في القسم السابق أيضاً أنه بمجرد اختيار هدف كبير بما فيه الكفاية لكي يكون تقريب توزيع باريتو المعمم دقيقاً فإننا نتوقع أن زيادة الحدود بمقدار Δ يترك ξ دون تغيير ولكن يزيد β بمقدار $\xi \Delta$. وبالتالي، فإننا نتوقع أن يكون متوسط الفائض في ذيل التوزيع خطي كدالة للهدف مع ميل $(1 - \beta)/\xi$ شريطة أن يكون $\xi < 1$.

وبالتالي، إذا قمنا بتعيين المتوسط الزائد لأهداف مختلفة، يجب أن نحصل على خط مستقيم على التوالي لقيم الحدود الكبير بما يكفي لتطبيق تقريب توزيع باريتو المعمم ولكن صغير بما فيه الكفاية أيضاً لتكون هناك نقاط كافية فوق الحدود لتقدير عينة من المتوسط ليكون دقيقاً. من المعقول تنفيذ إجراء التقدير بقيمة α التي تكون بالقرب من بداية هذا الجزء النهائي من الخط المستقيم إذا كان موجوداً.

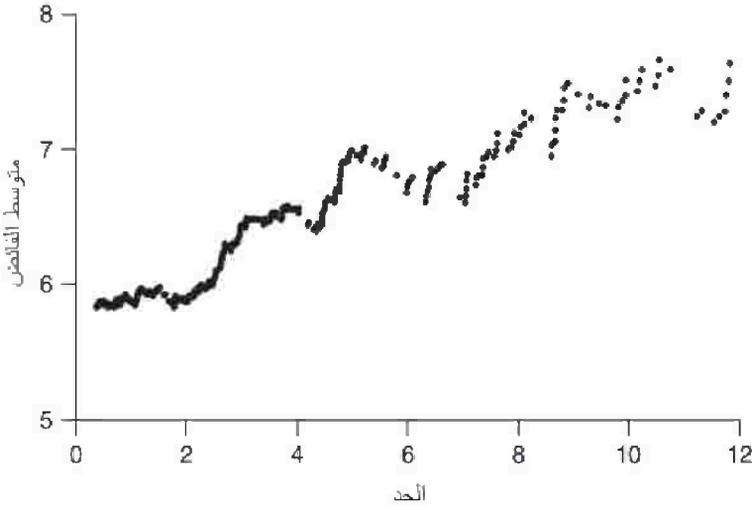
مثال 4-5: تقدير مقاييس المخاطر لموقع Amazon.com

لتوضيح ذلك، سننظر إلى بعض بيانات سعر السهم: يوضح الشكل 4-9 الخسارة



الشكل 4-9: الخسارة / المكاسب الأسبوعية لسهم شركة Amazon.com على مدار الفترة ما بين 1988 و2011

الأسبوعية (القيم السلبية تعبر عن المكاسب) لسعر سهم Amazon.com على مدى 14 عاما من بداية عام 1998 وحتى نهاية عام 2011. هناك نقطة من البيانات الأسبوعية التي نحسب من خلال نسبة 729 من الخسائر. من الواضح أن هناك تقلبات أعلى بكثير في الجزء الأول من هذه الفترة، ولكننا لن نحاول استخدام هذه التغيرات في التقلب ضمن إجراءات التقدير لدينا. ويبين الشكل 4-10 متوسط البيانات الزائدة مع اختلاف الحدود. يمكننا أن نرى أنه بالنسبة للحدود في نطاق من 0 إلى 2٪، فإن متوسط الفائض يستقر قرابة 6٪ خسارة، ولكن بالنسبة للحدود التي تعدت خسارة أكثر من 2٪، فإن متوسط الفائض يبدأ في الارتفاع. ويرتفع بطريقة غير منتظمة، ولكن هناك اتجاهًا تصاعديًا واضحًا. هذا يشير إلى أننا يمكن أن نستخدم قيمة خسارة 2٪ للحد u. ولكن أفضل خيار للحد يعد تعسفي إلى حد ما، وبالنسبة لهذه البيانات يمكن أن نأخذ أيضًا قيمة للحد قرابة الـ 6٪، والتي من شأنها أن تعطي فائض 125 (من التجاوزات).



الشكل 10-4: مخطط متوسط الفائض لبيانات سعر سهم Amazon

2-4-4 الخطوة الثانية تقدير المعلمات ξ و β

بعد تثبيت قيمة u ، فإن الخطوة الثانية في تلك العملية هي تقدير المعلمات ξ و β في توزيع باريتو المعمم الذي ينطبق فوق u . يمكننا القيام بذلك باستخدام نهج أقصى احتمال. تعد هذه طريقة قياسية إلى حد ما من الإحصاءات، ولكننا سنقدم ملخصاً عن كيفية عملها.

لنفترض أن لدينا M من القيم الملحوظة، ونتساءل عن احتمالية تلك القيم إذا كان التوزيع حقاً توزيع باريتو المعمم $G_{\xi, \beta}$ وتم اختيار كل ملاحظة بشكل مستقل؟ من الواضح أن احتمال الحصول على أي مجموعة محددة من القيم هو صفر ولكن يمكننا أن ننظر في الاحتمال المحدد على أنه الكثافة الاحتمالية للتوزيع العام للقيم الممكنة M التي تم تقييمها في مجموعة الملاحظات التي نلاحظها. طريقة أخرى للنظر في احتمال هو الحد من احتمال الحصول على نقاط في فترات صغيرة حول تلك التي نلاحظها حقاً، ولكن مع تطبيعها للسماح لتأثير إطارات الفاصل الزمني.

على سبيل المثال، لنفترض أن هناك ثلاث نقاط فقط فوق مستوى u والقيم التي تتجاوزها

u هي Y_1 و Y_2 و Y_3 . وبما أن كثافة توزيع باريتو المعمم هي $(1/\beta)(1 + \xi x/\beta)^{-1/\xi-1}$ فإن كثافة التوزيع المشترك بافتراض الاستقلالية هي نتاج هذا، مثال، الاحتمالية هي

$$(1/\beta^3)(1 + \xi Y_1/\beta)^{-1/\xi-1}(1 + \xi Y_2/\beta)^{-1/\xi-1}(1 + \xi Y_3/\beta)^{-1/\xi-1}.$$

الفكرة الآن هي البحث عن قيم اثنين من البارامترات التي تجعل هذه الصيغة على أكبر قدر ممكن. النتائج هنا صعب التعامل معها، خاصة عندما يكون لدينا أكثر من ثلاث ملاحظات. ولكن تعظيم الاحتمالات سوف ينتهي في نفس اختيار ξ و β كما يحدث مع تعظيم أي زيادة لدالة الاحتمال، مثل تربيع الاحتمال. دالة الزيادة التي تعمل بأفضل شكل هي تعظيم الاحتمالية اللوغاريتمية لتحويل النتيجة إلى مجموع. ولهذا نقوم بتعظيم

$$\log((1/\beta^3)(1 + \xi Y_1/\beta)^{-1/\xi-1}(1 + \xi Y_2/\beta)^{-1/\xi-1}(1 + \xi Y_3/\beta)^{-1/\xi-1}) \\ = -3 \log \beta + (-1/\xi - 1)(\log(1 + \xi Y_1/\beta) + \log(1 + \xi Y_2/\beta) + \log(1 + \xi Y_3/\beta)).$$

بوجه عام، مع وجود قيم فائضة لـ M التي هي Y_1, Y_2, \dots, Y_m يمكن أن نقدر ξ و β من خلال اختيار القيم ξ و β التي تعظم

$$-M \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^M \log \left(1 + \frac{\xi Y_i}{\beta}\right). \quad (8-4)$$

مراعاة القيود $\beta > 0$ و $1 + \xi Y_i/\beta > 0$ لكل قيم Y ويمكن إجراء هذا الحساب باستخدام Solver في جدول بيانات (كل ما تحتاجه هو تخمينات β و ξ ، عمود من القيم Y_i ، وعمود الذي يكون فيه اللوغاريتم $(1 + \xi Y_i/\beta)$ محسوب، ومن ثم خلية فيها الهدف المعطى بواسطة الصيغة (8-4)).

مثال 4-5: (تابع) تقدير مقاييس المخاطر لموقع Amazon.com

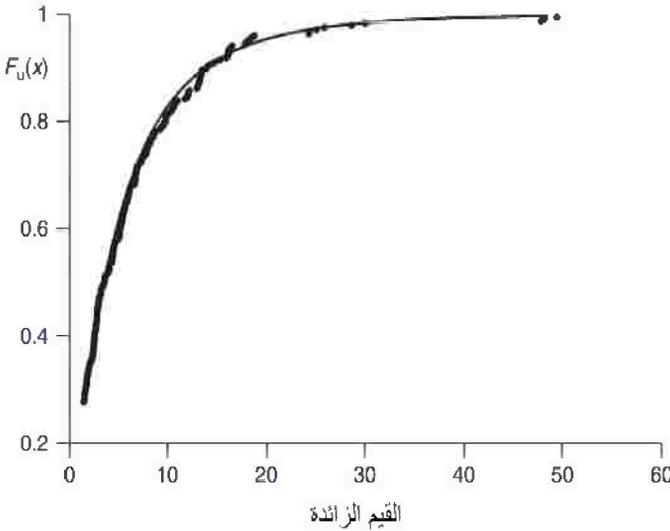
بالعودة إلى بيانات Amazon.com، لدينا 264 أسبوعا كانت الخسارة فيها أكبر من 1/2. يتم تعظيم تعبير لوغاريتمية الاحتمال عن طريق أخذ $\xi = 0.27$ و $\beta = 4.51$. يتوافق ذلك مع قيمة مؤشر ذيل $1/\xi = 3.7$. ويبين الشكل 4-11 مدى ملاءمة دالة التوزيع التراكمي الملائمة للبيانات لنقاط البيانات 264 فوق الحد. نعرض القيمة المقدرة لـ $F_u(x)$ كخط

والبيانات التجريبية كنقاط. على سبيل المثال، أكبر 10 خسائر هي 20.73 ويتم رسم ذلك بقيمة y من $264/254 = 0.962$.

3-4-4 الخطوة الثالثة. تقدير مقاييس المخاطر ذات الأهمية

الخطوة الثالثة والأخيرة هي استخدام التوزيع المجهز لتقدير القيم المهمة. سوف ننظر في تقدير كل من القيمة المعرضة للخطر والنقص المتوقع. ننظر أولاً في القيمة المعرضة للخطر. $\text{VaR}_\alpha(X)$ هي x مثل $F(x) = \alpha$. إذا كنا نعمل مع حد ثابت u و $x > u$ ثم

$$\begin{aligned} F(x) &= 1 - \Pr(X > x) \\ &= 1 - \Pr(X > u) \Pr(X > x \mid X > u) \\ &= 1 - (1 - F(u))(1 - F_u(x - u)). \end{aligned}$$



الشكل 11-4: توزيع باريتو المعمم المقدر مقارنة مع البيانات التجريبية للخسارة الأسبوعية % في موقع Amazon.com مع حد 2

بما أننا نفترض أن توزيع F_u هو $G_{\xi, \beta}$ مع قيم متوقعة $\hat{\xi}$ و $\hat{\beta}$ ، إذ أن $F(x) = \alpha$ تتطلب

$$\alpha = 1 - (1 - F(u))(1 + \hat{\xi}(x - u)/\hat{\beta})^{-1/\hat{\xi}}. \quad (9-4)$$

حل x يعطي القيمة المعرضة للخطر التي نريدها، ويمكننا إعادة ترتيب المعادلة (4.9) للحصول على x من حيث المعلمات الأخرى وحتى الحصول على التقدير التالي

$$\text{VaR}_\alpha(X) = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{1 - \alpha}{1 - F(u)} \right)^{-\hat{\xi}} - 1 \right).$$

العجز المتوقع عند مستوى α هو $ES_\alpha(X) = E(X|X > \text{VaR}_\alpha(X))$. لحساب ذلك يمكننا تحديد $\text{VaR}_\alpha(X)$ باستخدام حساباتنا السابقة في ذيل التوزيع لـ $v > u$ للتوزيع الزائد $F_v(x) = G_{\xi, \beta'}(x)$ حيث $\beta' = \beta + \xi(v - u)$ علاوة على ذلك، نعرف أن متوسط القيمة للتوزيع $G_{\xi, \beta}$ is $\beta/(1 - \xi)$ ولذا، العجز المتوقع هو

$$\begin{aligned} E(X|X > \text{VaR}_\alpha(X)) &= \text{VaR}_\alpha(X) + E(\text{excess over } \text{VaR}_\alpha(X)) \\ &= \text{VaR}_\alpha(X) + \frac{\beta_V}{1 - \xi}, \end{aligned}$$

حيث β_V هي القيمة المتوقعة β للمنحنى (الذيل) أعلى $\text{VaR}_\alpha(X)$ ، ولذا

$$\beta_V = \hat{\beta} + \hat{\xi}(\text{VaR}_\alpha(X) - u)$$

ولهذا

$$ES_\alpha(X) = \text{VaR}_\alpha(X) + \frac{\hat{\beta} + \hat{\xi}(\text{VaR}_\alpha(X) - u)}{1 - \hat{\xi}},$$

التي يمكن تبسيطها لإنتاج الصيغة النهائية

$$ES_\alpha(X) = \frac{\text{VaR}_\alpha(X) + \hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}}.$$

مثال 4-5: (يتبع) توقع إجراءات المخاطر لـ Amazon.com

يمكننا حساب القيمة المعرضة للخطر VaR بنسبة 99٪ وتراجع القيمة المتوقع لبيانات Amazon.com بالضبط بهذه الطريقة. نستخدم توقعاتنا السابقة لـ $\hat{\xi} = 0.27$ و $\hat{\beta} = 4.51$ بقيمة هدف $u = 2$. كما نحتاج أيضًا $1 - F(u)$ والتي هي احتمالية الخسارة بأكثر من u ، ونتوقع

ذلك من خلال نسبة الملاحظات فوق الهدف لـ 2، مثال $1 - F(u) = 264/729$ وهذا يعطي

$$\text{VaR}_{0.99}(X) = 2 + \frac{4.51}{0.27} \left(\left(\frac{0.01}{(264/729)} \right)^{-0.27} - 1 \right) = 29.3.$$

وبالتالي، يقدر التراجع المتوقع من خلال

$$\text{ES}_{0.99}(X) = \frac{29.3 + 4.51 - 0.27 \times 2}{1 - 0.27} = 45.6.$$

يمكننا مقارنة هذه الأشكال من خلال الملاحظات. من بين 729 ملاحظة نتوقع أن نرى 7 أعلى المستوى $\text{VaR}_{0.99}$: أعلى ملاحظة في المرتبة السابعة هي 27.86 والثامنة هي الأعلى عند 26.35. ولهذا فإن طريقة توزيع باريتو المعمم GPD تعطي تغاير أعلى بشكل طفيف بأكثر من المقترح من خلال النظر في أكبر خسائر فقط. قد يكون هذا محور نقاش سواء ما كان التوقع 29.3 أفضل من توقع معتمد على أعلى قيمة سابعة. بالرغم من ذلك، لا يوجد شك أنه كلما كان النهج أكثر تعقيداً كلما كان توقع التراجع أفضل، لأن هذا هو الوضع الذي يصبح فيه عنصر الاستقراء في التقدير أكثر أهمية. وتعطي البيانات التجريبية خسارة متوسطة بين نقاط البيانات السبع الأولى التي تبلغ 38.4 بينما توقع توزيع باريتو المعمم يعطي قيمة أعلى من القراءة 45.6.

ملاحظات

يحتوي هذا الفصل على أكثر التحليلات تعقيداً في هذا الكتاب، ولكن في نفس الوقت نفسه يغطي جزءاً كبيراً مما يمكن تناوله. وتوفر كتب Frey و McNeil و Embrechts مقدمة متكاملة للعديد من التقنيات في هذا الصدد، وهذه هي بداية جيدة لأي شخص يود معرفة المزيد.

قدمنا نموذج لتوقع مؤشرات المنحنيات ببداية هذا الفصل، وفي نهاية الفصل عرضنا نهج أكثر تعقيداً باستخدام الاحتمالات القصوى لتوقع $\hat{\alpha}$ (والتي هي عكس لمؤشر الذي (المنحنى)). ولكن هناك أكثر من نهج آخر: التقنية الأكثر شيوعاً تسمى بمتوقع Hills.

التوزيع الموضح في المثال 1.4 مرتبط جدًا بالتوزيع المعروف لـ Cauchy والذي له

دالة

$$f(x) = \frac{1}{\pi(1+x^2)}$$

والذي ليس له متوسط معروف بشكل جيد.

ترجع نظرية Fisher–Tippett إلى ورقة بحثية من هؤلاء المؤلفين لعام 1928، ولكن لم يكن هناك إثبات كامل لها حتى ظهور عمل لـ Gnedenko في 1943 وعادة ما يتم إضافة اسمه إلى اسم النتيجة، والتي أصبحت في النهاية Fisher–Tippett–Gnedenko. ولكن حتى أبحاث Gnedenko لم تجيب على كافة التساؤلات المرتبطة بهذه النتيجة واستمر العمل على النظرية بإثباتات جديدة حتى السبعينات.

المواد التي وصفناها حول التوزيعات الزائدة عادة ما تسمى بنظرية "القمم أعلى الهدف" (POT). وتعد هذه النسخة البارزة لنظرية القيم القصوى، وهي مفيدة أكثر من نظيرتها في التعامل مع العدد الأقصى من N من النماذج.

كتب Embrechts و Kluppelberg و Mikosch تغطي الأفكار النظرية الأساسية للقيم القصوى وتحتوي على دلائل للنظريتين الموضحتين. وتوفر الورقة البحثية لـ De Haan إثبات لنظرية القيم القصوى المعممة وتوفر تسلسل للسواكن التطبيقية $a_N = F^{-1}(1 - 1/N)$ و $b_N = 1/(Nf(a_N))$ واللاتان لهم ميزة البساطة التي تسمح باستخدامها في أمثلة. وتوفر هذه المراجع حالات أكثر دقة للتوزيع في $MDA(H0)$ (على سبيل المثال، سوف يتم ضمان ذلك في حال $\lim_{x \rightarrow \infty} f(x)(1 - F(x))/(f(x)^2) = -1$).

المراجع

- De Haan, L. (1976) Sample extremes: an elementary introduction. *Statistica Neerlandica*, 30, 161–172.
- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer.

- Fisher, R. and Tippett, L. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical roceedings of the Cambridge Philosophical Society*, 24, 180–190.
- Gnedenko, B. (1943) Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44, 423–453.
- McNeil, A., Frey, R. and Embrechts, P. (2005) *Quantitative Risk anagement*. Princeton University Press.

تمارين

1-4 مقارنة المنحنيات العلوية والسفلية لسعر الصرف

الشكل 2-4 تم إنشاؤه باستخدام البيانات في جدول البيانات Exchange Rate Data. BRMch4- xlsx. استخدم هذه البيانات للقيام بتحليلات مشابهة للتحركات في سعر الصرف في الاتجاه المعاكس (التراجع في قيمة الدولار الأمريكي مقابل الجنيه الإسترليني). عل تعتقد أن مؤشر المنحنى قرب الـ 5 ملائم أيضًا لهذه الحالة؟

2-4 توزيع فريشيه

تعرب الشركة عن قلقها إزاء الدعاية السيئة المحتملة الناجمة عن ضمان على موقعها على الإنترنت ("سنقوم بإصلاح المشكلة المتعلقة بجهاز الإرسال "الراوتر" في أقل من 20 دقيقة أو سنقدم لك خدمة الإنترنت مجاناً لمدة عام). افترض أن المدير يبحث في البيانات في 20 يوم عمل لتقدير الوقت الأقصى الذي تستغرق المشكلة ذات الصلة بجهاز الإرسال. في كل يوم كان هناك ما بين 25 و30 استفسار من قبل العملاء من هذا النوع.

الحد الأقصى من بين هذه الأوقات يختلف من يوم لآخر ولكن متوسط قيمة 20 مرة مختلفة هي 12 دقيقة. نفترض أن الوقت المطلوب يمثل مؤشر المنحنى ذو قيم عالية من $\alpha = 5$ ، وبالتالي تحديد توزيع الحد الأقصى اليومي لعدد المرات. استخدام هذا التوزيع لتقدير احتمال أن يتم إلغاء الضمان في يوم معين، وبالتالي العدد المتوقع من الأيام قبل حدوث ذلك. [ستحتاج إلى استخدام صيغة لمتوسط توزيع فريشيه. إذا كانت $F(x) = 1 - (x - \alpha)^{-\alpha}$ فإن

المتوسط يعطى بواسطة دالة غاما $(1 - \alpha)^{-1}$. هنا نحتاج $\Gamma(0.8) = 1.164$.

3-4 سلوك المنحنى في خليط من المعدلات

لنفترض أننا نضع نموذج تكلفة الذهب في وقت محدد في المستقبل (10 يناير 2020) على النحو المعطى من خلال التوزيع العادي بمتوسط μ والانحراف المعياري σ . فكرتنا هي أنه سيكون هناك قيمة متوسطة للذهب في عام 2020 ولكن السعر سوف يتذبذب حول تلك القيمة. نحن لا نعرف أي من هذه الأرقام سوف يتحقق، ولكننا نعتقد أن μ سيكون حوالي \$ 1500 لكل أوقية، ونحن نعتقد أن التقلب اليومي، الذي يقاس بـ σ ، سيكون حوالي \$ 100. وبشكل أدق، فإننا نمثل عدم اليقين حول μ بالقول أن μ يتم استخلاصها من التوزيع العادي بمتوسط 1500 والانحراف المعياري 100 ونمثل عدم اليقين حول σ بالقول أن σ مستمد من توزيع طبيعي بمتوسط 100 وانحراف معياري 20 وتنتج هذه العملية ما يسمى بخليط من التوزيعات العادية.

استخدم نموذج جدول البيانات BRMch4-MixtureOfNormals.xlsx لاستكشاف الطريقة التي يؤثر بها خليط التوزيعات العادية على سلوك المنحنى.

4-4 حساب متوسط القيم بتوزيع باريتو المعمم

أظهر إذا كان التوزيع له كثافة $f(x) = (1/\beta)(1 + \xi x/\beta)^{-1/\xi-1}$ ثم متوسطه $\beta/(1 - \xi)$. (ملحوظة: استخدم التكامل غير المعرف

$$\int (x/\beta)(1 + \xi x/\beta)^{-(1/\xi)-1} dx = \frac{x + \beta}{\xi - 1} (1 + \xi x/\beta)^{-(1/\xi)}$$

وضع في اعتبارك تكامل من 0 إلى R ثم اترك الحد العلوي R يتجه إلى ما لا نهاية.

5-4 مخطط المتوسط عندما لا يكون المتوسط محدد

خلق بعض البيانات من توزيع بقيمة $\xi = 1.2$ باستخدام الصيغة الخلية $=1/(1-RAND())^1.2$ في جدول بيانات (مع مؤشر منحنى منخفض لا يوجد متوسط توزيع). تحقق مما يحدث لمتوسط المخطط الزائد في هذه الحالة. يجب ملاحظة أنه من الغريب

أنه سيسير بشكل جيد بخط مستقيم في أغلب النطاق. هل يمكنك شرح ما يجري؟ (وهذا يدل على قيمة التحقق من تناسب الحصول عليها من أقصى تقدير متوقع بالطريقة التي تتم في الشكل 4-11).

4-6 تقدير البارامترات من متوسط الأرقام الزائدة

ينظر محلل في بيانات متعلقة بتكاليف الرسوم من إنهاء الأعمال التجارية بعد إفلاسها. ولديه بيانات عن 900 حدث من هذا القبيل ويحسب متوسط الأرقام الزائدة التي تستخدم أهداف 10 ملايين دولار و20 مليون دولار. وهناك 50 حدثًا تزيد تكاليفها الإجمالية عن 10 ملايين دولار، حيث يبلغ متوسط التكاليف 50 دولارًا قدره 19 مليون دولار (أي بزيادة قدرها 9 ملايين دولار)، ويوجد 15 حدثًا بتكاليف إجمالية تتجاوز 20 مليون دولار، بمتوسط حسابي (15 مليون دولار) (بزيادة قدرها 12 مليون دولار). تقدير قيم β و γ لتوزيع باريتو المعمم للفائض فوق 25 مليون دولار.

4-7 تقدير القيمة المعرضة للمخاطر وقيمة العجز المتوقع باستخدام نظرية القيمة القصوى

استخدم العملية الموضحة لبيانات سهم أمازون لتقدير القيمة المعرضة للخطر 99٪ وقيمة العجز المتوقع 99٪ للخسائر اليومية على مؤشر S & P 500. البيانات مدرجة في نموذج جدول البيانات BRMch4-S & P500.xlsx.