مبرهنة قانون الأعداد الكبيرة: إذا كانت $X_1, X_2, \dots, X_n$ متتالية من المتغيرات العشوائية المستقلة والتي لها جميعاً نفس التوزيع بمتوسط μ وتباين σ^2 محدودين عندئذٍ من أجل كل $\varepsilon > 0$ يكون

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) = 0$$

الإثبات: لدينا (حسب خواص التوقع)

$$E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \mu$$

ولدينا أيضاً حسب خواص التباين:

$$Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

وبتطبيق متراجحة تشيبتشيف نجد:

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) \le \frac{Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\varepsilon^2} = \frac{\sigma^2/n}{\varepsilon^2}$$

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) \le \frac{\sigma^2}{n\varepsilon^2}$$

وبأخذ نهاية الطرفين نجد:

$$0 \le \lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) \le \lim_{n\to\infty}\frac{\sigma^2}{n\varepsilon^2} = 0$$

$$\Rightarrow \lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| - \varepsilon\right) = 0$$

ملاحظة: لقانون الأعداد الكبيرة تطبيقات هامة في الإحصاء وبخاصة إذا كنا حيال تقدير متوسط التوزيع الذي تتبعه متتالية متغيرات عشوائية $(X_n)_{n\ge 1}$ والقانون يشير إلى أنه كلما كان n عدد المتغيرات المستقلة كبيراً كان

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

أقرب إلى μ.

تقريب التوزيع الثنائي (الحداني) إلى التوزيع الطبيعي :

إذا كان X متغيراً عشوائياً له التوزيع الثنائي بوسيطين n و p فإن X هو عبارة عن مجموع متغيرات عشوائية $X_1, X_2, \dots, X_n$ برنولية لكل منها الوسيط p. و بحسب مبرهنة النهاية المركزية فإن التوزيع التقريبي لـ X في حالة n كبيرة كفاية هو التوزيع الطبيعي بمتوسط np وتباين npq أي أن :

$$X = \sum_{i=1}^{n} X_i \sim N(np, npq) \quad ; \quad q = 1-p$$

وهكذا يمكننا من جديد استخدام جدول التوزيع الطبيعي المعياري لحساب احتمالات تتعلق بالمتغير العشوائي الحداني (الثنائي) و لكن بصورة تقريبية بشرط تحقق الشرطين

$np \geq 5$ و $nq \geq 5$ ، وعندها يكون

$$p(x_1 \leq X \leq x_2) = p(x_1 - 1/2 \leq Y \leq x_2 + 1/2)$$

حيث $Y \sim N(np, npq)$

حيث أضفنا وطرحنا $\frac{1}{2}$ كعملية تصحيح من أجل الاستمرار ($\frac{1}{2}$ يدعى عامل التصحيح)

تمرين :

إذا قذفنا قطعة نقود متوازنة (10) مرات فما هو احتمال أن نحصل على الصورة ثلاث مرات أو أربع مرات أو خمس مرات

الحل: بفرض X متغير عشوائي يدل على ظهور الصورة عند رمي قطعة النقود 10 مرات فيكون لـ X التوزيع الحداني بالوسيطين $n=10$ و $p=0.5$ ويكون لـ X دالة الكثافة الاحتمالية

$$f_X(x) = C_x^n p^x q^{n-x} \quad ; \quad x = 0, 1, \dots, 10$$

$$f_X(x) = C_x^{10} (0.5)^x (0.5)^{10-x} = C_x^{10} (0.5)^{10} \quad ; \quad x = 0, 1, \dots, 10$$

ومنه

$$p(3 \leq X \leq 5) = p(X=3) + p(X=4) + p(X=5)$$

$$= C_3^{10} (0.5)^{10} + C_4^{10} (0.5)^{10} + C_5^{10} (0.5)^{10}$$

$$= (0.5)^{10} [120 + 210 + 252] = 0.5683$$

طريقة ثانية: بملاحظة أن $np = 10(0.5) = 5$

$$nq = 10(0.5) = 5$$

حسب مبرهنة النهاية المركزية يكون $p(3 \leq X \leq 5) = p(3 - \frac{1}{2} \leq Y \leq 5 + \frac{1}{2})$

حيث $Y \sim N(np, npq) = N(5, 2.5)$

ومنه:

$$P(3 \le X \le 5) = P(2.5 \le Y \le 5.5)$$

$$= P\left(\frac{2.5-\mu}{\sigma} \le \frac{Y-\mu}{\sigma} \le \frac{5.5-\mu}{\sigma}\right)$$

$$= P\left(\frac{2.5-5}{\sqrt{2.5}} \le Z \le \frac{5.5-5}{\sqrt{2.5}}\right)$$

$$= P(-1.58 \le Z \le 0.316)$$

$$= \phi(0,316) - \phi(-1.58)$$

$$= \phi(0,316) - [1-\phi(1.58)]$$

$$= 0.6255 - 0.0571 = 0.5684$$

بالمقارنة بين طريقتي الحل نلاحظ أن القيمة الناتجة صحيحة إلى ثلاث أرقام عشرية على الرغم من أن $n=10$ فقط.

عزوم العينة ودوالها

* تعريف العينة العشوائية

نقول عن مجموعة المتغيرات العشوائية $X_1, X_2, \ldots, X_n$ إنها عينة عشوائية من الحجم n للمتغير العشوائي X إذا كانت مستقلة ولها جميعاً قانون (تابع) توزيع X أي أن:

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

* تعريف الإحصاءة:

هو كل دالة في عينة عشوائية لا تتعلق بوسطاء مجهولة فإذا كانت $X_1, \ldots, X_n$ عينة عشوائية في متغير عشوائي X فإن الدوال

$$X_5 + X_6 \,,\; \sum_{i=1}^{n} X_i^2 \,,\; \prod_{i=1}^{n} X_i \,,\; \frac{1}{n}\sum_{i=1}^{n} X_i$$

تكون جميعها إحصاءات بينما الدوال

$$\sum_{i=1}^{n} (X_i-\mu)^2 \,,\; \prod_{i=1}^{n} X_i - \sigma_i^2 \,,\; \sum_{i=1}^{n} X_i^2 - \mu\sigma$$

ليست إحصاءات طالما μ و σ مجهولين، ولكنها تغدو إحصاءات عندما يكون الوسيطان μ و σ معلومين.

* تعريف متوسط العينة: إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من الحجم n للمتغير العشوائي X عندئذٍ متوسط العينة بالتعريف هو الإحصاء

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

* تعريف تباين العينة:

إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من الحجم n للمتغير العشوائي X عندئذٍ تباين العينة بالتعريف هو الإحصاء

$$(*) \ldots S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

برهان (*)

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2)$$

$$= \sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + n\overline{X}^2$$

ونعلم أنّ

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$ وبالتعويض نجد

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{2}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i + n\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{2}{n}\left[\sum_{i=1}^{n} X_i\right]^2 + \frac{1}{n}\left[\sum_{i=1}^{n} X_i\right]^2$$

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} X_i\right]^2$$

وبتوحيد المقامات في الطرف الثاني

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n}$$

وبتقسيم الطرفين على $n-1$ نجد:

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

وهو المطلوب #

انتهت المحاضرة التاسعة