محاضرات الإحصاء الرياضي

فهرس المحتويات

مقدمة

هذه المطبوعة

هذه المطبوعة هي عبارة عن محاضرات الإحصاء حسب البرنامج الوزاري لمقياس "إحصاء 2" للسنة الثانية علوم التسيير. برمج هذا المقياس لطلبة السنة الثانية، لكي يستفيد الطلبة من القاعدة التي اكتسبوها عند دراسة الإحصاء الوصفي في السنة الأولى، لكن هدفه الأساسي هو التمهيد لدراسة الإحصاء التطبيقي في السنة الثالثة. هدف هذا المقياس هو تقديم علم الإحصاء الرياضي، أي الأساس الرياضي للإحصاء التطبيقي.

باعتبارها فرعا من الرياضيات، تدرس مادة هذا المقياس في كليات العلوم و الهندسة، لكن تقديم هذه المادة لطلبة العلوم الإنسانية يتضمن صعوبة خاصة. هذه المطبوعة هي ثمرة تجربة سنوات عدية في تدريس الإحصاء بكلية العلوم الاقتصادية لجامعة المسيلة، ولقد حاولنا أن نستفيد من هذه التجربة لصياغة محتوى المقياس بطريقة تلائم مستوى طلبة هذه الكلية و طبيعة التخصص. لتحقيق هذا الغرض حرصنا على ربط المفاهيم والقواعد النظرية باستخداماتها التطبيقية؛ فعملنا على إعطاء أمثلة محلولة عن كل مفهوم جديد. و لأن فهم القواعد الرياضية يكون أسهل إذا كان للمتلقي خلفية عن المشكلة التي يحتاج حلها إلى استخدام هذه القواعد، عملنا في كثير من الأحيان إلى التقديم لبعض الدروس أو النظريات بمسألة تكون بمثابة التمهيد، وأحيانا بمثابة مشكلة ننطلق منها لنتوصل إلى النظرية. هذا وننبه طلبتنا الأعزاء إلى أنه يفترض بهم عند دراسة الإحصاء الرياضي أن يكونوا قادرين على استيعاب المفاهيم الرياضية بعموميتها و لا يبقوا خيالهم حبيس الأمثلة والمسائل المعطاة، فالإحصاء الرياضي غير الإحصاء التطبيقي الذي يعنى بتطبيقات هذه المفاهيم فيما بعد.

يتضمن البرنامج المقرر على ثمانية فصول، أطولها الفصل الثاني المعنون "المتغيرات العشوائية". من أجل الموازنة بين الفصول رأينا أن نعيد تجزئة محتويات البرنامج. فأعدنا تقسيم محتويات الفصل الثاني إلى 3 فصول نظرا لحجمه، وجمعنا محتويات الفصل السابع و الثامن في فصل واحد لتعلقهما بموضوع واحد. ولقد قسمنا الفصول إلى مباحث، بحيث يوافق المبحث محاضرة واحدة في أغلب الأحيان، و التزمنا في الغالب الأعم بالمنهج المقرر، لكن سوف يجد القارئ أننا توسعنا في بعض الجوانب من خلال الملحقات، فله أن يلم بهذه الاستطرادات إن رأى أنه قد تمكن من فهم النقاط الرئيسية المقررة، و إلا فإننا ننصحه بأن يمر عليها مرور الكرام. و غني عن الذكر أن محتوى هذه المطبوعة من نظريات وقواعد ليس من إبداع مؤلفها، و إنما هي قواعد مبسوطة في المراجع جمعناها وعرضناها بأسلوب رأينا أنه الأنسب لمستوى طالب كلية العلوم الاقتصادية. وإذ نقدم لطلبتنا و زملائنا هذا العمل المتواضع، نهيب بهم أن لا يبخلوا علينا بملاحظاتهم وتعليقاتهم حتى نستفيد منها لطبعات مقبلة بحول الله.

متطلبات المقياس

فيما يتعلق بما تحتاجه متابعة وفهم هذا المقياس، من المهم التمييز بين الفصل الأول وبقية الفصول الأخرى. فالفصل الأول الذي يتضمن المفاهيم الأساسية لعلم الاحتمالات لا يحتاج استيعابه إلى مستوى عالي في الرياضيات، أما باقي الفصول فيتطلب فهمها أن يقوم الطالب بمراجعة عدد من المفاهيم الرياضية أغلبها متضمنة في برنامج الرياضيات للسنة الأولى. تتمثل هذه المفاهيم أساسا في الدوال، الاشتقاق، التكامل (خاصة التكامل بالتجزئة) والدوال الأسية. كما يحتاج الطالب إلى قاعدة بسيطة في مفاهيم اللوغاريتم، التكامل الثنائي والسلاسل الشهيرة.

كلمة إلى الطلبة

كثيرا ما نلاحظ أن الطلبة يستخدمون التمارين المقدمة في السلاسل كنماذج أو شبه قوانين في حد ذاتها يحاولون حفظها بينما هي في الحقيقة مجرد وسيلة لفهم الدرس. هذا التشبث بالشكل دون المضمون في محاولة يائسة لمواجهة الامتحان دون

فهم حقيق لمضمون المادة هو نتيجة حتمية بالنسبة لمن لا يتابع المحاضرات والتطبيقات بالمراجعة المستمرة و الفورية. وحسب رأينا فإن الصعوبة التي يواجهها الطلبة في هذا المقياس سببها أنه مقياس يعتمد أساسا على الفهم أكثر مما يعتمد على التذكر. وهذا الفهم لا يتأتى عن طريق التلقي من الأستاذ، مهما بذل هذا الأخير ومهما كانت مهارته، وإنما يحتاج إلى جهد مستقل يبذله الطالب بمفرده مع قدر من التركيز و المثابرة. "الوصفة السحرية" لفهم هذه المادة، هي المراجعة بجرعات منتظمة و فورية (بعد كل محاضرة قبل النوم[1]!) مع شيء من التركيز على القواعد والمفاهيم حتى يتم فهمها فهما جيدا. وليعمل الطالب على تعميق فهمه من خلال تمارين السلاسل و لكن لا يتخذها "نماذج" جامدة أو قواعد إضافية. إن هدف الأستاذ والجامعة ككل هو إعداد الطالب لمواجهة المشكلات المعقدة للتسيير، وهذا الهدف لا يتحقق إلا بتنمية الذكاء والتزود بعدد من التقنيات المساعدة. إن الجائزة الحقيقية التي يجب أن يتوقعها الطالب من دراسة بالجامعة هي تكوين قدرة على التعلم الذاتي أكثر من تجميع كم من المعارف التي قد لا يحتاجها أبدا، و هي من جهة أخرى، تكوين ذهنية مستقلة قادرة على تحليل المشكلات والوضعيات المعقدة وصياغتها في شكل واضح ودقيق ومن ثم إبداع حلول لها من خلال تفكيره الخاص. هذه القدرة لا تتأتى إذا عود الطالب نفسه على إعمال فكره مطولا في المسائل التي تطرحها التمارين مما يعطي الطالب القدرة على التحليل والتركيب والاستنباط والاستدلال كأسس التفكير المنتج والمبدع. إن الوصول إلى هذه القدرة على مواجهة مشكلات وحلها هي غاية أساسية للتعليم الجامعي وهي أحسن رأسمال يجمعه الطالب ليستثمره حياته العامة والخاصة معا.

نبذة تاريخية عن تطور علم الاحصاء[2]

> "وهكذا فإن البحث يتقدم عبر مراحل منفصلة و مستمرة من الحدس، التعصب، الإثارة و الحمى. و ذات يوم تتحقق أخيرا الفرحة و يتذوق طعمها من عاش تلك اللحظات الفريدة. [...]   ألبرت أينشتاين[3]

قبل الشروع في دراسة الطرائق المختلفة للإحصاء الرياضي يستحسن أن يحيط الطالب بنظرة عن التطور التاريخي للإحصاء كممارسة وكعلم، وأن يطلع على مجموعة من أبرز من كتبوا في هذا العلم.

الفترة ما قبل الميلاد إلى غاية القرن 18: تدل الحفريات التي وجدت في أماكن متعددة على استخدام الإحصاء من قبل عدد من الحضارات القديمة عبر المعمورة. منذ القدم استخدم الحكام والأمراء الإحصاء كوسيلة للرقابة، و أداة لإدارة المملكة أو المدينة أو المقاطعة، واستخدموا في ذلك تعداد السكان وجرد السلع والموارد المختلفة. في الحضارة السومرية، التي سادت في بلاد ما بين النهرين 5 آلاف إلى ألفي سنة قبل الميلاد، والتي ازدهرت فيها التجارة بشكل كبير، كانت قوائم من السلع والأشخاص تدون على ألواح من الصلصال، وقد وجدت حفريات مشابهة تثبت استخدام الجرد في عهد الحضارة المصرية التي سادت 3 آلاف سنة قبل الميلاد. الحضارة المصرية التي قامت على التسيير والتقسيم الدقيق لمياه النيل اتسمت إدارتها بالمركزية الشديدة وهذا الذي أعطى الأهمية **للتدوين** كوسيلة للمراقبة، فقد كان للمصريين القدامى مدارس يتعلم فيها الموظفون القراءة والكتابة والقوانين المعمول بها، وكان مما يتعلمه الموظف أن لا اعتبار لأمر أو عقد ما لم يكن مكتوبا. واستخدم الجرد لدى جميع الحضارات القديمة تقريبا كالحضارة الصينية والهندية واليابانية واليونانية والرومانية، وكذا

---

[1] يعلمنا علم نفس التربية أن أكثر من 60 بالمئة من المعلومات التي نتعلمها ننساها في التسع ساعات الأولى. فأنقذ معلوماتك في الساعات الأولى قبل أن تتبخر!

[2] أخذت معظم المعلومات التاريخية عن: جون جاك دروزبيك، أساسيات في الإحصاء، سلسلة (SMA)، دار (Ellips)، 1996، ص 2.

[3] من كتابه: كيف أرى العالم. ترجمه إلى الفرنسية ريجي هونريو، باريس، فلاماريون، 1979. عن:
Microsoft ® Encarta ® 2006. © 1993-2005 Microsoft Corporation.

حضارة الإنكا في الساحل الغربي لأمريكا الجنوبية (ابتداءا من القرن 12 إلى غاية 1572) . في هذا العهد كان الإحصاء عبارة عن جرد المواد والأفراد وأحيانا نجد نظاما لتصنيف المعلومات لكن لم يوجد دليل على عمليات معالجة لهذه المعطيات.

في العهد الإسلامي كان الخليفة عثمان (ر) أول من أمر بالتدوين لإحصاء المستفيدين من عطايا بيت المال، أما في أوربا فنجد أن أول الآثار عن عمليات التعداد ترجع إلى 1086 فقط وبالتحديد في بريطانيا. أما في فرنسا فإن عمليات التعداد ترجع إلى القرن 14 الذي شهد ميلاد أول تسجيلات عقود الحالة المدنية وإجبارية تسجيل عقود الازدياد في عهد فرنسوا الأول. في فرنسا دائما تجدر الإشارة إلى أنه في القرن17 حين أراد "كولبيرت" - أب الإدارة الفرنسية – أن يدفع ببلاده إلى المستوى الصناعي الذي بلغته بريطانيا في ذلك الوقت، أسس إدارة مركزية قوية... وكان من منجزاته أن شهدت وزارته (1630-1660) عددا من عمليات التحقيق الكبرى. وشهدت ألمانيا وبريطانيا تطورا مشابها بالإضافة إلى دول أخرى. وقد كان "قرانت" (1620- 1674 : GRANT) أول من استعمل في 1662 مصطلحات علم السكان مثل الخصوبة وطول مدة الحياة؛ كما قارن بين معدلات ولادة الإناث والذكور. وقد طور هذا العالم مع عالم آخر هو بيتي (PETTY) طريقة لتعداد السكان من خلال المعطيات الثانوية (عن عدد المساكن، عدد الوفيات...) تدعى "طريقة المضاعف" (Multiplicateur) عرفت بعد ذلك تحسينات متتالية على أيدي علماء آخرين منهم خاصة "لابلاس" (LAPLACE)في 1785.

ظهور نظرية الاحتمالات في قرن17 و18: تاريخيا ارتبط ظهور نظرية الاحتمالات بألعاب الحظ التي كانت سائدة بكثرة في أوربا في القرن السابع عشر وتنظمها البنوك بشكل خاص. لكن قلة انتشار طباعة الكتب والأجواء الدينية السائدة التي لا تبارك هذه الألعاب منعت انتشار الكتابات في هذا الشأن. وينسب البعض أول الكتابات في علم الاحتمالات إلى العالم "باسكال" (1623 -1662 PASCAL) الذي كتب عما أسماه آنذاك "هندسة الحظ" (La géométrie du hasard). وكان ذلك من خلال رسائل له مع زميله المعروف هو الآخر "فرمات" (1601 – 1665: FERMAT) . وتذكر في هذا الصدد بشكل خاص المسألة التي طرحها على باسكال أحد هواة الألعاب "كم ينبغي من رمية لمكعبي نرد حتى يمكن المراهنة بتفاؤل على الحصول على مجموع 12؟ ". ثم جاء علماء آخرون كانت لهم إضافات بارزة في هذه الفترة مثل هايجان (1695 – 1629 : HUYGEN)، جاك برنولي (JACQUES BERNOULLI)، موافر (MOIVRE) وكذا لايبنيتز (1646 – 1716: LEIBNIZ). كما ساهم في هذه الفترة التي سبقت القرن 19 علماء كبار أمثال (GAUSSE, BAYES, LAPLACE) عرفت نظرية الاحتمالات على أيديهم إنجازات كبيرة.

القرن 19: في هذا القرن برزت إحدى أهم عناصر نظرية الاحتمالات وهي "التوزيع الطبيعي" وذلك لقياس نسبة الخطأ في مجال الحسابات الفلكية. كان هذا من ثمرة عمل العالمين لابلاس وقوس(GAUSSE وLAPLACE). في هذا القرن أيضا ظهرت حسابات الارتباط لقالتو (GALTOU) كما برزت أسماء مثل كتلت (QUETLET) وآخرون .

القرن العشرون: نظرية الاحتمالات كما نراها الآن، أي بصياغة رياضية ناضجة في شكل قوانين مبرهن عليها رياضيا، إنما تبلورت في القرن العشرين وبالضبط في بدايته. ومن الأسماء التي برزت في الفترة الأولى (1980 – 1920) من هذا القرن نجد من بريطانيا بيرسون (KARLE PEARSON) ومن روسيا ماركوف (MARKOV) ومن فرنسا بوريل (BOREL). في الفترة الثانية (1921 – 1932) درست مسائل التوقع، حيث كان لفيشر (FISHER) دورا بارزا.

في الفترة الممتدة من 1933 إلى نهاية الحرب العالمية الثانية برزت اختبارات الفروض على يد نايمان (NEYMAN) وإيقون بيرسون (EGON PEARSON) وبداية النظرية الحديثة للمعاينة لنايمان (NEYMAN) بالإضافة إلى خطط التجارب لفيشر. بداية من الخمسينات تكاثرت الكتابات في مجال الإحصاء حيث عرفت نظرية التقدير وتحليل البيانات. وبالتدريج انتشر استخدام الإحصاء في الميادين المختلفة والعلوم التجريبية والإنسانية.

ملخص: تاريخيا إذا كانت أولى استعمالات الإحصاء ارتبطت بحاجة الدولة لتنظيم الجباية والتجنيد ودراسة السكان فإن أولى الدراسات في حساب الاحتمالات (أصل الإحصاء الرياضي) ارتبطت أول الأمر بمسائل ألعاب الصدفة والحظ كمجال جديد أثار فضول عدد من العلماء الذين أسسوا هذا العلم في القرن 17. التطور السريع لعلم الاحتمالات كفرع من الرياضيات كان في بداية القرن 20 لكن أهم عناصر الإحصاء الرياضي كما هو معروف الآن تبلورت في النصف الأخير منه.

تعريف علم الإحصاء

تعريف مصطفى الخواجة [1] : "يقصد بعلم الإحصاء، الطريقة الإحصائية، وهي تلك الطريقة التي تمكن من:

- جمع الحقائق عن الظواهر المختلفة في شكل قياسي،
- تسجيل بيانات تلك الحقائق في جداول تلخيصية،
- عرض بيانات تلك الجداول بيانيا وتحليلها بهدف معرفة اتجاهات هذه الظواهر والعلاقات فيما بينها.

أي أن علم الإحصاء يختص بالطريقة العملية لجمع وتنظيم وتلخيص وعرض وتحليل البيانات بهدف الوصول إلى نتائج مقبولة وقرارات على ضوء هذا التحليل. أي يمكن القول بإيجاز شديد أنه" علم استنباط الحقائق من الأرقام بأسلوب علمي وبطريقة علمية."

... "ويمكن تقسيم علم الإحصاء إلى قسمين وهما الإحصاء الوصفي والإحصاء التحليلي, وفرع الإحصاء الذي يهدف فقط إلى وصف وتحليل مجموعة معينة دون الوصول إلى نتائج أو استدلال خاص بالمجموعات الأكبر أو الأخرى فإنه يسمى بالإحصاء الوصفي، أما الإحصاء التحليلي فيهتم بعمليات التنبؤ والتقدير عن طريق استخدام جزء من المجموعة للوصول إلى قرار أو حكم عام يمكن تطبيقه على المجموعة كلها، ولذلك يعتمد في جزء كبير منه على نظرية الاحتمالات."

تعريف جلاطو جيلالي[2] : "الإحصاء هو علم جمع وترتيب معلومات خاصة بظاهرة معينة وقياس الوقائع كأساس للاستقراء ".

تعريف جون جاك دراوزبيك[3] : يعرف هذا العالم الإحصاء بأنه ذلك العلم الذي يشمل " مجموعة الطرق التي تهدف إلى معالجة المعطيات ..." ويقول عن موضوع علم الإحصاء: "يتعلق الأمر بمعرفة كيفية الحصول على تلك البيانات، معرفة من أين يتم تجميعها وبأي شكل يكون ذلك التجميع."

تعريف دومنيك سلفاتور[1] : " الإحصاء هو مجموع الطرق الرياضية المتعلقة بجمع، وعرض، وتحليل، واستخدام المعطيات الرقمية. هذه العمليات تمكن من استخلاص استنتاجات واتخاذ قرارات إزاء حالة عدم التأكد التي نواجهها في مجال الاقتصاد ومجال الأعمال أو في علوم اجتماعية وفيزيائية أخرى."

---

[1] مصطفى الخواجة: مقدمة في الإحصاء، الدار الجامعية، الإسكندرية، 2002، ص 2.

[2] جلاطو جيلالي، الإحصاء مع تمارين ومسائل محلولة، 2002، ديوان المطبوعات الجامعية، ص 3.

[3] دروزبيك، مرجع سابق.

ويواصل الكاتب « نميز بين الإحصاء الوصفي والإحصاء الاستدلالي (Statistique Inductive) الأول يلخص، يحوصل ويحلل كما من من المعطيات، أما الثاني فيسقط على الكل من خلال دراسة الجزء، الكل يسمى في هذه الحالة المجتمع (أو العالم Univers) والجزء يسمى العينة. صحة الإسقاط تتطلب إذا أن تكون العينة ممثلة وأن تكون احتمال الخطأ محسوبا." فيما يخصنا، يهتم هذا المقياس بدراسة الفرع الثاني من الإحصاء المتمثل في الإحصاء الاستدلالي، وهناك من يسمي هذا الفرع من الإحصاء "الإحصاء التطبيقي". من المهم ذكر هذه التسميات حتى يعلم الطالب أن بإمكانه البحث عن مادة المقياس في مراجع تحت هذه العناوين وغيرها مثل الاقتصاد القياسي، الإحصاء الاقتصادي، الاحتمالات، الاحتمالات والمتغيرات العشوائية أو ببساطة الإحصاء.

الاحصاء التطبيقي

تعريف "ريجينالد لافوا"[2] : "المسألة الأساسية للإحصاء التطبيقي تتمثل نظريا كما يلي: نريد دراسة عدد من الخصائص (كالعمر، الوزن، التوجه السياسي...) لمجتمع ما، لكن لأسباب مختلفة لا يمكن أن نشمل بالدراسة كل أفراد المجتمع. لهذا نلجأ إلى دراسة جزء من المجتمع (عينة) لدراسة هذه الخصائص، وعند إتمام دراسة العينة نعمل على تعميم على المجتمع ككل الحقائق المشاهدة مع التقييم، لفرض عدم الخطأ في هذا التعميم".

حوصلة: بصفة عامة ومن خلال جميع التعريفات السابقة يمكن القول أن علم الإحصاء يهتم بكيفية جمع وترتيب وعرض البيانات وكذا كيفية تحليلها للخروج بخلاصة مفهومة.

[1] دومينيك سالفاتور، الاقتصاد القياسي والإحصاء التطبيقي، سلسلة شوم، دار ماك قراو هيل، 1985، ص 1.

[2] ريجينالد لافوا، الإحصاء التطبيقي، الكيبيك، 1981، ص 1.1

الفصل I. تذكير بالمفاهيم الأساسية للاحتمالات

مفاهيم أساسية التـرميز

من بين علوم الرياضيات العليا يعتبر البعض الاحتمالات على أنها الأكثر تعقيدا و "الأكثر علوا" !!، والحقيقة غير ذلك. إنها لا تعدو أن تكون بالنسبة لمن يريد حقا فهمها لعبة مسلية تتلخص في بضعة قواعد بديهية. ولا يضاهي بساطة الاحتمالات إلا تعدد استخداماتها وتواجدها في جميع الميادين، ما يفسر حتمية دراستها على جميع الشعب تقريبا. بالنسبة لعالم الاقتصاد والتسيير فإن فهم حساب الاحتمالات هو أداة يومية لمعالجة المشاكل المطروحة واتخاذ القرار. فقرارات المسير، بل وحتى رب البيت، تبنى في 99 % من الحالات على معلومات غير مؤكدة.

المبحث 1. مفاهيم أساسية

مفهوم التجربة، الحدث والاحتمال
خصائص الاحتمال
القواعد الأساسية في حساب الاحتمال
تعريف باسكال للاحتمال

1 مفهوم التجربة، الحدث والاحتمال Epreuve, événement, probabilité

(أ) الاحتمال و الحدث Evénement et probabilité

كثيرا ما يخلط الطلبة بين هذين المفهومين لارتباطهما ببعض. فالحدث العشوائي هو واقعة أو نتيجة ما، أما الاحتمال فهو عدد بين الصفر والواحد يعبر عن حظوظ وقوع الحدث (ليس شرطا أن يكون زمن وقوع الحدث هو المستقبل، فقد يكون الماضي أو الحاضر). سئل الرئيس العراقي السابق – قبل حرب الخليج الأولى– ما هو احتمال انهزامكم في هذه الحرب ؟ فأجاب: "واحد إلى مليون". عندما نرغب في التعبير بشكل دقيق على مدى إمكانية وقوع حدث معين فإننا عادة نستعمل عبارات مثل: 100% للحدث المؤكد أو 50% للحدث المحتمل و1‰% مثلا للحدث المستبعد، إذن نحن نستخدم الكسور في سلم تصاعدي من 0 إلى1، بحيث يرمز 0 للاستحالة و1 للتأكد.

مثال. احتمال الحصول على صورة عند رمي قطعة نقدية هو 2/1، و احتمال الحصول على الوجه "6" عند رمي حجر نرد هو 6/1 .

يمكن جمع الأعداد أو طرحها و يمكن أن تخضع للجداء و القسمة أما عمليات التقاطع و الاتحاد، .. فهي عمليات على المجموعات و ليست على الأعداد. من أجل ذلك لا يصح أن نكتب احتمال تقاطع (أو اتحاد) احتمال. تمثل هذه القاعدة الذهبية الأولى في الاحتمالات وفي هذا المقياس ككل.

ويجب التمييز بين الاحتمال والإمكانية (الإمكانية هي حدث). فالاحتمال في مفهوم العلم هو عدد يقيس حظوظ وقوع شيء ما نسميه نتيجة أو حدث أو إمكانية. أما الإمكانية فهي حدث أو نتيجة ما من بين أحداث أو نتائج أخرى. يختلف عن هذا المفهوم العلمي تعريف الناس للاحتمال. فكثيرا ما تطلق كلمة الاحتمال و يقصد بها إمكانية، فيقال مثلا

"إن هذا احتمال ممكن" و الصحيح إن هذه إمكانية واردة" أو يقال "إذا رمينا حجر نرد هناك 6 إحتمالات" و الصحيح " هناك 6 إمكانيات أو 6 نتائج محتملة"، ...

(ب) التجربة Epreuve

لشرح المفهوم المجرد للتجربة و تمييزها عن الحدث يمكن القول أن التجربة هي أم الحدث أو أم النتيجة. لأن التجربة تتفرع بالضرورة إلى أحداث. ففي المقولة السابقة، التجربة هي الحرب بينما الهزيمة هي نتيجة ممكنة للحرب.و التجربة قد تقبل نتيجتين أو أكثر.

ومفهوم التجربة في علم الاحتمالات مفهوم عام و مرن، فإذا كنا ندرس احتمال الحصول على الوجه 6 عند رمي قطعة نرد تكون التجربة هي الرمي، و إذا كنا ندرس احتمال عدد معين من الوحدات التالفة لآلة ما يمكن اعتبار كل وحدة منتجة كتجربة، وإذا كنا ندرس احتمال عدد معين من الطلبة الراسبين في مقياس ما نعتبر كل طالب كتجربة... نقول احتمال حدث أو احتمال نتيجة ولا نقول إحتمال تجربة.

2 خصائص الإحتمال

عادة ما نعبر عن هذه الخصائص بالطريقة التالية:

- الاحتمال هو عدد موجب تماما أو معدوم (لا يكون سالبا).
- مجموع احتمالات أحداث تجربة ما يساوي الواحد.

ويمكن إضافة خاصية ثالثة تستنتج بديهيا من الخاصيتين السالفتين وهي أن الاحتمال يكون محصورا بين 0 و 1. أي أنه لا يمكن أن يكون سالبا ولا أن يكون أكبر من الواحد.

3 الأركان الخمسة في حساب الاحتمالات

هناك خمس قواعد أساسية في حساب الاحتمال نذكرها الآن باقتضاب لإبراز أهميتها ونعود لشرحها فما بعد وسنحتاج إلى استخدام هذه القواعد في جميع فصول المقياس.

1. احتمال وقوع حدث يساوي 1 مطروحا منه احتمال الحدث المعاكس. مجموع احتمال الحدث واحتمال الحدث المعاكس يساوي 1.
2. احتمال وقوع حدثان "أ" و "ب" يساوي احتمال وقوع الأول مضروبا في احتمال وقوع الثاني لما يكون الأول قد وقع فعلا.
3. احتمال وقوع حدثان مستقلان يساوي جداء الاحتمالين أي احتمال الحدث الأول مضروبا في احتمال الحدث الثاني.
4. احتمال وقوع الحدث وعكسه يساوي الصفر، ونقول أن الحدثان متنافيان.
5. احتمال وقوع حدث "ا" أو "ب" يساوي جمع احتمالي الحدثين مطروحا منه احتمال تحققهما معا.

4 القاعدة السادسة أو حساب الاحتمال حسب تعريف باسكال للاحتمال

عرف بليز باسكال (Blaise Pascal : 1623) الاحتمال بالشكل التالي:

"احتمال حدث هو عدد الحالات الملائمة لوقوع الحدث مقسوما على عدد الحالات الممكنة، إذا اقترضنا أن كل الحالات لها نفس الاحتمال في الوقوع." [1]

مثال: ما هو احتمال الحصول على عدد زوجي عند رمي قطعة نرد ؟ بين كل من التجربة، الحدث والاحتمال في هذا المثال.

الجواب: هناك ثلاث حالات ملائمة للحصول على عدد زوجي (2، 4 و6). أما العدد الكلي للحالات الممكنة فهو 6: (1، 2، 3، 4، 5، 6). **وبافتراض أن كل الحالات الممكنة لها نفس الاحتمال** فإن احتمال الحصول على عدد زوجي هو $\frac{3}{6}=\frac{1}{2}$.

تنبيه: لا يمكن استخدام هذه العلاقة إذا لم تكن احتمالات الحالات متساوية.

مثال2. صندوق به 7 كريات منها 5 حمراء. نسحب 3 كريات معا. ما هو احتمال أن تكون كلها حمراء؟ بين كل من التجربة والحدث في هذا المثال.

عدد الحالات الملائمة C_5^3 وعدد الحالات الممكنة: C_7^3. إذا الاحتمال هو $\frac{C_5^3}{C_7^3}=\frac{10}{35}$. التجربة هي السحب من الصندوق، الحدث أو النتيجة هي الحصول على ...

مثال3. فوج مكون من 10 طلبة. نسحب بالقرعة اسم من العشرة. ما هو احتمال أن يكون الطالب أحمد؟ بين كل من التجربة والحدث.

نسحب (بدون إعادة) عينة من 3 أسماء من العشرة. ما هو احتمال أن يكون منهم الطالب أحمد؟

الجواب: 1) احتمال الحدث الأول أو النتيجة الأولى هي $\frac{1}{10}$،

2) عدد الطرق الممكنة للعينة: C_{10}^3، عدد الحالات الملائمة لكي يكون أحمد في العينة:

$$C_{10-1}^{3-1}=C_9^2 \qquad C_n^x=\frac{n!}{(n-x)!x!} \Rightarrow C_9^2=36$$

الاحتمال هو إذا: $\frac{C_9^2}{C_{10}^3}=\frac{36}{10(9)(8)/3}=\frac{36}{240}=\frac{3}{20}$

التجربة هي السحب، النتيجة أو الحدث هي أن يكون الطالب أحمد، ...

مثال 4. يتنافس أحمد مع 3 زملائه على أعلى نقطة في كل من الامتحانات الستة للسداسي. إذا كانت حظوظ الطلبة الأربعة متساوية، ما هو احتمال: أن يفوز أحمد بأعلى نقطة في كل من الامتحانات الستة؟ أن يفوز أحد الطلبة (أيا كان) بأعلى نقطة في الامتحانات الستة؟

الجواب:

1) هناك $6^4 = 4096$ حالة ممكنة لنتائج المنافسة، منها حالة فوز أحمد بجميع المقاييس؛ إذا الاحتمال هو 1/4096.

2) هناك 4 طلبة إذا هناك 4 حالات لفوز أحد الطلبة بجميع المقاييس، إذا الاحتمال هو 4/4096 .

5 خلاصة

الاحتمال هو عدد لا يزيد عن 1 و لا يقل عن 0.

[1] ريجينالد لافوا 1981، ص 2 و3.

التجربة والحدث والاحتمال هي مفاهيم لا يجب الخلط بينها. التجربة يتولد عنها أحداث (نتائج أو حالات) مختلفة. التجربة مفهوم مرن يتطلب أحيانا نظرة ذكية وخيال. من المهم اكتساب هذه المهارة في تحديد ما هي التجربة أو التجارب في مسألة ما لأن ذلك هو المفتاح لفهم و حل المسألة.

هناك خمس قواعد في حساب الاحتمال هي الأركان الأساسية لعلم الاحتمالات. هذه القواعد متعلقة ب:

- احتمال الحدث المعاكس،
- باحتمال تحقق حدثين معا،
- باحتمال تحقق حدثين معا إذا كانا مستقلان،
- باحتمال تحقق أحد حدثين،
- و متعلقة باحتمال تحقق الحدث و عكسه معا.

المبحث 2. الترميز أو التعبير الرياضي عن الاحتمالات

"الطبيعة هي كتاب لغته الرياضيات" جاليلي (1564-1642)

استخدام نضرية المجموعات

التعبير الرياضي عن قواعد جمع وضرب الاحتمالات

نظرية بايز

نستخدم الترميز من أجل التوصل إلى تعبير دقيق و واضح لقواعد الحساب الاحتمالي وهي ذاتها القواعد الأربعة المذكورة في الجزء الأول.

نعبر عن احتمال حدث ما بطريقة رياضية فنكتب $P(A)$

ونعبر عن احتمال وقوع الحدث : $X = x$ كما يلي: $P(X = x)$ أو $P(x)$.

مثال: احتمال الحدث: "الحصول على الوجه 5" عند إلقاء حجر نرد يكتب: $P(X = 5) = 1/6$ ، أو باختصار: $P(5) = 1/6$

و أحيانا نختصر أكثر فنكتب: $P = 1/6$.

1 استخدام نظرية المجموعات للتعبير عن الأحداث العشوائية

من خلال البنود التالية تستخدم نظرية المجموعات للتعبير عن الأحداث العشوائية:

1. نعبر عن النتائج الممكنة لتجربة ما ب Ω، وتسمى المجموعة الكلية أو فضاء العينة.
2. نعبر عن الحدث بمجوعة جزئية A من فضاء العينة، حيث A هي مجموعة من النتائج الممكنة للتجربة.
3. إذا انتهت التجربة بنتيجة تمثل عنصرا من A نقول أن الحدث A قد تحقق.
4. الحدث الذي يحتوي على نقطة أو عنصر واحد من Ω يسمى عادة حدث بسيط.

مثال. لتكن لجينا تجربة هي إلقاء مكعب نرد. أكتب مجموعة فضاء العينة ثم عبر عمليا عن الأحداث التالية:

الحدث A: الحصول على العدد 6 (حدث بسيط) $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{6\}$

الحدث B: الحصول على عدد زوجي $B = \{2, 4, 6\}$

الحدث C: الحصول على عدد أولي $C = \{2, 3, 5\}$

الحدث D: الحصول على عدد فردي $D = \{1, 3, 5\}$

مثال 2: لتكن لدينا تجربة هي رمي قطعتين نقديتين على التوالي: أكتب مجموعة فضاء العينة ثم عبر عن:

الحدث A : الحصول على مرتين كتابة (حدث بسيط) $\Omega = \{PP, PF, FP, FF\}$, $A = \{PP\}$

الحدث B : الحصول على كتابة مرة واحدة $B = \{FP, PF\}$

الحدث C : الحصول على كتابة في الرمية الأولى $C = \{PF, PP\}$

5. من بين الأحداث الممكنة في تجربة ما أيا كانت، الحدث Φ يمثل الحدث المستحيل لأنه لا يمكن أن يتحقق عنصر منها. $P(\Phi) = 0$.
6. من بين الأحداث الممكنة في تجربة ما أيا كانت، حدث المجموعة الأساسية Ω نفسها، وهو الحدث الأكيد لأنه لابد أن يتحقق أحد عناصرها على الأقل. $P(\Omega) = 1$
7. بتطبيق عمليات مثل الإتحاد والتقاطع، الطرح، الجمع .... على المجموعات نحصل على مجموعات جديدة جزئية من Ω ومن ثم أحداث جديدة في Ω . من ذلك :

$A \cup B$ هو الحدث: إما A أو B أو كلاهما.

$A \cap B$ هو الحدث: A و B في وقت معا.

C_A هو الحدث المعاكس ل A.

$A - B$ هو الحدث: A لكن ليس B .

8. إذا كان $A \cap B = \Phi$ نقول أن A و B متنافيان (أو غير متلائمان) أي لا يمكن وقوعهما معا (mutuellement exclusifs).

مثال: نرمي قطعة نقدية مرتين: إذ كان A هو الحدث "مرتين كتابة" و B "صورة على الأقل".

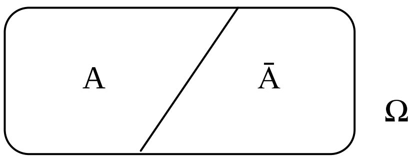
$A = \{PP\}$. $B = \{PF, FP, FF\}$ $\Phi = A \cap B$

2 التعبير الرياضي عن قواعد حساب الاحتمالات

أ) الحدث المعاكس أو التعبير الرياضي عن القاعدة رقم 1. Evénement contraire

نعبر عن الحدث المعاكس ل A ب $\bar{A}$ أو A' واحتماله هو احتمال عدم تحقق الحدث A، ونكتب :

$$\mathbf{P(\bar{A}) + P(A) = 1 \iff P(\bar{A}) = 1 - P(A)}$$



مثال: نرمي قطعة نقدية ونرمز ب P للكتابة و F للصورة (الوجه). نلاحظ أن:

$$P(P) = P(F') = 1 - P(F) \iff P(P) + P(F) = 1$$

مثال 2: عند رمي حجر نرد فإن احتمال الحصول على العدد 5 هو: $P(5) = 1/6$ ، فما هو الحدث المعاكس في هذه الحالة وما احتماله؟

الحدث المعاكس هو الحصول على عدد غير 5، واحتماله هو: $P(5') = 1 - P(5) = 1 - (1/6) = 5/6$.

مثال 3: نرمي حجر نرد، ما هو احتمال الحصول على عدد زوجي، ماهو الحدث المعاكس وما هو احتماله؟

$$P(\text{nombre pair}) = P(2 \text{ ou } 4 \text{ ou } 6) = 3/6$$

الحدث المعاكس هو الحصول على عدد غير زوجي، و احتماله:

$$P(\text{impair}) = 1 - P(\text{pair}) = 1 - (3/6) = 3/6.$$

(ب) احتمال وقوع الحدث "A" و "B" أيا كانت (قاعدة رقم 2).

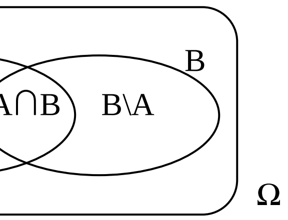
$$P(A \cap B) = P(A) * P(B/A)$$

$$P(A \cap B \cap C) = P(A) * P(B/A) * P(C/(A \cap B))$$

A, B, C أحداث ما (مستقلة أو لا، متنافية أو لا) ، $P(B/A)$ يسمى الاحتمال الشرطي لـB علما أن A محقق.

ومن المعادلة الأولى نحصل على:

لما $P(A) > 0$ $\quad P(B/A) = P(A \cap B) / P(A)$



رسم1 الحدث B/A غير الحدث B\A

حيث A تصبح فضاء المعاينة بما أن A محقق.

مثال: 1) أحسب عند إلقاء حجر نرد احتمال الحصول على قيم أقل من 4 (حدثB).

2) أحسب احتمال الحصول على نتيجة أقل من 4 إذا علمت أن الوجه المحصل لمكعب النرد عدد فردي (حدثA).

3) أحسب احتمال الحصول على قيمة أكبر أو يساوي 4 إذا علمت أن النتيجة عدد فردي.

$$P(B) = P(1 \text{ou } 2 \text{ ou } 3) = P(1) + P(2) + P(3) = 3/6$$

$$P(B/A) = P(B \cap A)/P(A)$$

$$P(B \cap A) = P(\text{impaire et } \leq 4) = P(1 \text{ ou } 3) = P(1) + P(3) = 1/6 + 1/6 = 2/6$$

$$P(B/A) = P(B \cap A)/P(A) = 2/6 \ / \ 3/6 = 2/3$$

(ج) احتمال وقوع الحدث "A" و "B" لما "A" و "B" مستقلان (قاعدة رقم 3).

$$\mathbf{(P(B/A) = P(B)) \qquad P(A \cap B) = P(A) * P(B)}$$

وهو تعريف استقلال حدثين، أي أن وقوع B لا يتأثر بوقوع A أو عدم وقوعه نقول أن A وB مستقلان،

$$P(C/(A \cap B)) = P(C) \qquad P(A \cap B \cap C) = P(A) * P(B) * P(C)$$

مثال: نرمي حجر نرد وقطعة نقدية معا. ما هو احتمال الحصول على الصورة والعدد 6 ؟ (نتيجة مكعب النرد مستقلة عن نتيجة القطعة النقدية).

$$P(A \cap B) = P(A).P(B) = 0.5 * 1/6 = 1/12.$$

مثال 2. نلقي قطعة نقدية مرتين. أحسب احتمال الحصول على صورة في الرمية الأولى وفي الرمية الثانية.

$$P(FF) = P(A \cap B) = P(A) P(B) = 0.5 * 0.5 = 0.25$$

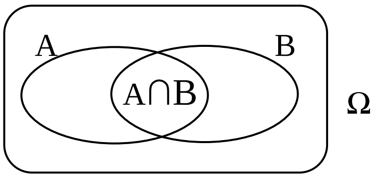
مثال 3. صندوق به 5 كريات 2 حمراء و3 بيضاء. نسحب كرية نسجل لونها ثم نعيدها للصندوق ونكرر العملية 3 مرات.

- أحسب احتمال الحصول على 2 كريات حمراء، 3 كريات حمراء (أحداث مستقلة).
- كيف يكون الاحتمال في حالة كون السحب بدون إرجاع الكرية (أحداث غير مستقلة)؟

$$P(RR) = P(R_1 \cap R_2) = P(R_1)\,P(R_2) = 2/5 * 2/5 = 8/25$$
$$P(RRR) = P(R_1 \cap R_2 \cap R_3) = P(R_1)\,P(R_2)\,P(R_3) = 2/5 * 2/5 * 2/5 = 8/125$$
$$P(RR) = P(R_1 \cap R_2) = P(R_1)\,P(R_2/R_1) = 2/5 * 1/4 = 2/20$$
$$P(RRR) = P(R_1 \cap R_2 \cap R_3) = P(R_1)\,P(R_2/R_1)\,P(R_3/(R_1 \cap R_2)) = 2/5 * 1/4 * 0 = 0$$

(د) **احتمال وقوع حدث "A" أو "B"** (**القاعدة رقم 4**).

$$\mathbf{P(AUB) = P(A) + P(B) - P(A \cap B)}$$

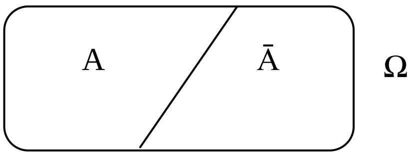


(هـ) **احتمال وقوع حدث "أ" أو "ب" لما "أ" و "ب" متنافيان** (**القاعدة رقم 5**).

لتكن الأحداث المتنافية A, B

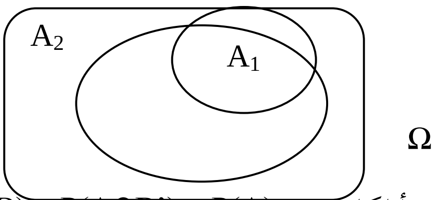
$$\mathbf{P(AUB) = P(A) + P(B) \qquad (P(A \cap B) = 0)}$$

$$P(AUBUC) = P(A) + P(B) + P(C) \qquad (P(A \cap B \cap C) = 0)$$

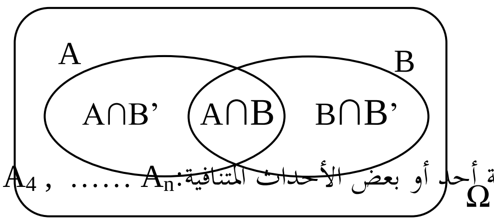


(و) **قواعد إضافية مهمة**

- من أجل $A_1 \subset A_2$ فإن: $P(A_1) \le P(A_2)$ و $P(A_2 - A_1) = P(A_2) - P(A_1)$

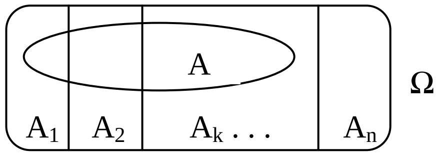


- من أجل A وB أحداث أيا كانت: $P(A \cap B) + P(A \cap B') = P(A)$



- إذا كان A هو نتيجة أحد أو بعض الأحداث المتنافية: $A_1, A_2, A_3, A_4, \ldots\ldots A_n$

$$P(A) = P(A \cap A_1) + P(A \cap A_2) + P(A \cap A_3) + \ldots\ldots + P(A \cap A_n)$$

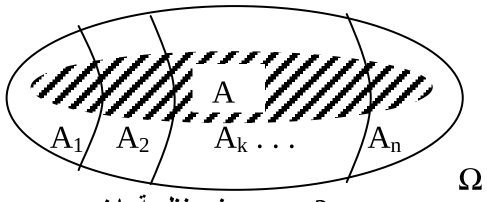


3 نظرية الاحتمال السببي أو نظرية بايز Théorème ou règle de BAYES

لتكن A_1، A_2، A_3، . . . A_k، . . . ، A_n أحداث متنافية فيما بينها حيث اتحادها يشكل المجموعة الكلية (الأساسية) Ω، وA حدث ما يتحقق عن طريق واحد أو أكثر من الأحداث A_k، إذا علمنا أن A تحقق، نحسب احتمال تحققه عن طريق الحدث A_k كما يلي:

$$P(A_k / A) = \frac{P(A_k)P(A/A_k)}{\sum_{k=1}^{n} P(A_k)P(A/A_k)} = \frac{P(A \cap A_k)}{P(A)}$$

تسمى هذه النظرية نظرية الاحتمال السببي لأنها تمكن من حساب احتمال أن يكون حدث ما (A_k) هو المسبب لوقوع حدث آخر (A).



رسم 2 رسم يوضح نظرية بايز

مثال: وظفت أمينة مكتب (A_1) بمكتب للمحاسبة حيث تولت طبع 20 % من الفواتير. يشغل المكتب عاملتين أخريين إحداهما (A_2) تطبع 30% من الفواتير والأخرى (A_3) 50%. ترتكب الموظفة الجديدة أخطاء في 5% من الفواتير، بينما نسبة الخطأ لدى الثانية (A_2) 2% ولدى الثالثة (A_3) 1%.

أخذت فاتورة بشكل عشوائي فتبين أن بها أخطاء. استبعدت الأولى أن تكون هي من أنجزت الفاتورة بحجة أنها لا تنجز إلا 20% من الفواتير، وردت عليها العاملات الأخريات بأن نسبة الأخطاء لديها هي الأكبر(5%).

1. أحسب احتمال أن تكون الموظفة الجديدة (A_1) هي التي حررت الفاتورة وقارن مع احتمال أن يكون مصدر الخطأ هو A_2 أو A_3.
2. أحسب مجموع الاحتمالات الثلاث.
3. أحسب احتمال أن تكون فاتورة مختارة عشوائيا من مجموع المراسلات، أن تكون بها أخطاء.

$$P(A_1 / A) = \frac{P(A_1)P(A/A_1)}{\sum_{k=1}^{3} P(A_k)P(A/A_k)} = \frac{0.2*0.05}{(0.2*0.05)+(0.3*0.02)+(0.5*0.01)} = 0.238$$

$$P(A_2 / A) = \frac{P(A_2)P(A/A_2)}{\sum_{k=1}^{3} P(A_k)P(A/A_k)} = \frac{0.3*0.02}{(0.2*0.05)+(0.3*0.02)+(0.5*0.01)} = 0.2857$$

$$P(A_3 / A) = \frac{P(A_3)P(A/A_3)}{\sum_{k=1}^{3} P(A_k)P(A/A_k)} = \frac{0.5*0.01}{(0.2*0.05)+(0.3*0.02)+(0.5*0.01)} = 0.476$$

يظهر من الحساب أن الاحتمال الأكبر هو أن تكون A_3 هي التي حررت الفاتورة.

2. مجموع الاحتمالات $P(A_1/A) + P(A_2/A) + P(A_3/A) = 1$ لأنها تمثل احتمالات الأحداث المتنافية الثلاث.

3. احتمال وجود خطأ في مراسلة ما:

$$P(A) = \sum P(A_k)P(A/A_k) = (0.2*0.005)+(0.3*0.02)+(0.5*0.01) = 0.012$$

4 خلاصة

باستخدام نظرية المجموعات كأساس للترميز في مجال الاحتمالات يمكن الحصول على صياغة أكثر دقة للمفاهيم المختلفة. بهذه الطريقة نستخدم:

رمز التقاطع $\cap$ بدلا عن عبارة "و" (et) مثال: احتمال "الوجه 2 و 5" في رميتي نرد: $P("2"\cap"5") = P(2) * P(5)$

رمز الإتحاد U بدلا عن عبارة" أو" "ou" مثال: احتمال الوجه 2 أو 5 في رمية نرد $P("5"U"2") = P(5) + P(2)$

رمز المتمم C_A أو $\bar{A}$ بدلا عن عبارة "عكس A "؛ $P(\bar{A}) = 1 – P(A)$

من خلال هذا الترميز يمكن أن نعبر بسهولة عن القواعد الخمسة الأساسية لحساب الاحتمالات.

- $P(A\cap B) = P(A)* P(B/A)$

$\Rightarrow P(B/A) = P(A\cap B) / P(A)$ (بشرط $P(A) > 0$)

- عندما يكون الحدثان مستقلان $P(A\cap B) = P(A)* P(B)$

- $P(AUB) = P(A) + P(B) – P(A\cap B)$

- عندما يكون الحدثان متنافيان أي $(P(A\cap B) = 0)$: $P(AUB) = P(A) + P(B)$

مسألة. نرمي قطعة نقدية مرتين، نسمي A "مرتين كتابة" وB "كتابة في المرة الأولى"، عبر عن الحدث:

A ،B، $\bar{A}$، $A\cap B$ ، AUB ، A –B ، B – A

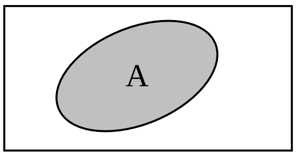
$A = \{PP\}, B = \{PP, FP\}, \bar{A} = \{PF, FP, FF\},$

$A\cap B = \{PP\}, AUB = \{PP, FP\}, A – B = \Phi, \quad B – A= \{FP\}$

5 ملحق

(أ) التعبير الهندسي عن الاحتمالات

قبل الشروع في حل مسألة مركبة للاحتمالات يستحسن تحليلها باستعمال أشكال هندسية توضح عناصر المسألة (الأحداث) والعلاقات بينها. يستخدم لهذا الغرض شجرة الاحتمال (أنظر الملحق) ومخطط فين. تبين شجرة الاحتمال الأحداث المتنافية التي تنتج عن التجربة الواحدة أو المكررة وذلك من خلال أغصان تتفرع من أصل، أما مخطط فين فيستخدم لتمثيل الأحداث الفرعية دوائر داخل مستطيل يمثل التجربة.

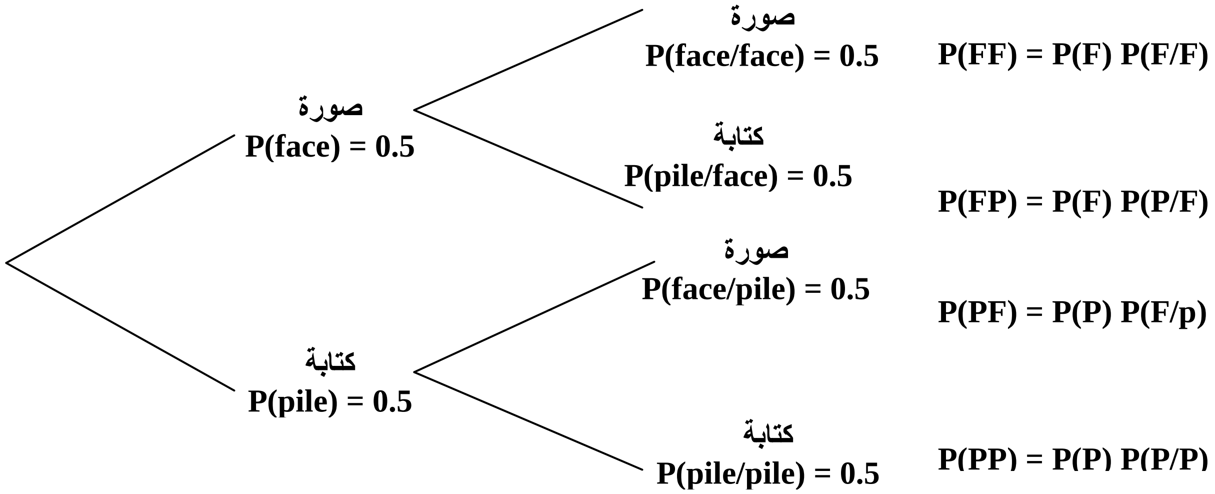


رسم1 مخطط فين

يراعى في رسم الشجرة أن يكون مجموع احتمالات كل تفريعة يساوي الواحد. التفريعة هي بمثابة شجرة فرعية تحتوي أحداث متنافية، لكونها تمثل النتائج المحتملة لتجربة جزئية.

مثال. نرمي قطعة نقدية مرتين. أحسب احتمال الحصول على مرتين صورة.

$P(\text{face} \cap \text{face}) = P(\text{face}) * P(\text{face/face}) = 0.5 * 0.5 = 0.25.$



ب) في مفهوم الحدث العشوائي

يجب الملاحظة أن كلمة حدث عشوائي لا تعني أن الحدث لا يخضع لأي قانون، بل المقصود أننا نتحدث عن حدث لا نعلم مسبقا ما إذا كان سيقع أو لا يقع. الهزيمة التي وقعت في الحرب وأي هزيمة كانت لها أسبابها وليست محض مصادفة عمياء. والحقيقة أن لا شيء في الطبيعة يقع بالمصادفة. فلا معنى لكلمة مصادفة إلا أننا لم نقصد وقوع الشيء. فعندما أقول التقيت بفلان صدفة، فهذا يعني أنني لم أقصد ولم أخطط لمقابلته. لكن هناك أسباب أدت إلى هذه الملاقاة منها أنني سلكت طريقا معينا ... كذلك إذا رمينا مكعب نرد 6 مرات فإننا لا نعلم إذا كنا سنحصل على مرة واحدة الوجه (5). لذلك قيمة $P(X)$ (مثلا $P(5) = 1/6$) هي قيمة نظرية، لكن إذا رمينا مكعب عدد كبير جدا من المرات (1000 مرة مثلا) فنتوقع أن عدد مرات الحصول على الوجه (5) سيكون قريبا جدا من العدد 1000/6. موضوع علم الاحتمالات هو البحث في قوانين الأحداث العشوائية، ولذلك أطلق عليه اسم "هندسة الحظ".

ج) في حساب عدد الحالات الممكنة أو الملائمة

بالإضافة إلى التوفيقات والترتيبات والأس، نحتاج أحيانا لحساب عدد الطرق الممكنة أو الملائمة إلى مفهوم surjection.

$$\text{surj}(n, k) = k\,[\text{surj}(n-1, k) + \text{surj}(n-1, k-1)] \quad , \quad (n, k > 0),$$
$$\text{Surj}(n, 1) = 1 \,,\ \text{Surj}(1, 1) = 1,\ \text{surj}(1, k > 1) = 0$$

مثلا: في المثال (4) أحسب احتمال أن يفوز كل طالب على الأقل بمقياس واحد:

عدد الحالات الملائمة هو :

$$\text{Surj}(6, 4) = 4[\text{surj}(5, 4) + \text{surj}(5, 3)]$$

لحساب ذلك نحتاج إلى حساب سلسلة من القيم:

$\text{Surj}(5, 3) = 3[\text{surj}(4, 4) + \text{surj}(4, 2)]$, mais: $\text{Surj}(4, 2) = 2[\text{surj}(3, 2) + \text{surj}(3, 1)]$,
mais: $\text{Surj}(3, 2) = 2[\text{surj}(2, 2) + \text{surj}(2, 1)]$,
mais : $\text{Surj}(2, 2) = 2[\text{surj}(1, 2) + \text{surj}(1, 1)] = 2(0+1) = 2$
$\Rightarrow \text{Surj}(3, 2) = 2(2+1) = 6, \ldots$

n k	2	3	4	5	6
2	2	0	0	0	0
3	6	6	0	0	0
4	14	36	24	0	0
5	30	150	240	120	0
6	62	540	**1560**	1800	720

الاحتمال هو إذا 1560/4096 .

الفصل II. المتغيرة العشوائية

مفهوم المتغيرة العشوائية المتقطعة
مفهوم المتغيرة العشوائية المستمرة و توزيعها الاحتمالي

المبحث 1. مفهوم المتغيرة العشوائية المتقطعة وتوزيعها الاحتمالي[1]

مفهوم المتغيرة العشوائية
مفهوم المتغيرة العشوائية المتقطعة
التوزيع الاحتمالي للمتغيرة العشوائية المتقطعة
شروط دالة الكثافة للمتغيرة العشوائية المتقطعة
التمثيل البياني لدالة الكثافة للمتغيرة العشوائية المتقطعة
دالة التوزيع للمتغيرة العشوائية المتقطعة

مسألة: أجريت دراسة على 1000 طفل أصيب خلال السنوات الثلاث الأولى من عمره بمرض ما.

بينت الدراسة أن احتمال الإصابة مرتبط بالزمن (X: السنة) من خلال دالة الكثافة التالية:

$$f(x)=\begin{cases} cx^2, & 0<x<3 \\ 0 & \text{sinon} \end{cases}$$

- ✓ أحسب احتمال أن تكون إصابة طفل مختار عشوائيا من العينة المدروسة في السنة الأولى.

يعالج المرض لمدة شهر، شهر ونصف، أو 3 أشهر حسب الجدول التالي:

X الأشهر	1	1.5	3
الاحتمال	0.5	0.3	0.2

- ✓ أحسب احتمال أن تكون مدة علاج طفل من العينة شهر ونصف على الأكثر.

1 مفهوم المتغيرة العشوائية

هي قيمة متغيرة يلحق بقيمها احتمالات تحقق كل قيمة. يرمز للمتغيرة ع بحرف لاتيني كبير. ونميز بين م ع المتقطعة وم العشوائبة المتصلة أو المستمرة.

مثال : في تجربة إلقاء مكعب نرد يمكن أن نسمي الوجه الذي يستقر عليه الكعب متغيرة عشوائية X. القيم الممكنة لX هي: 1، 2، 3، 4، 5، 6. بكل قيمة يمكن أن نلحق احتمال تحققها، وهو هنا 1/6. ونكتب مثلا :

$$P(X = 1) = f(1) = 1/6,\ \ P(X = 2) = f(2) = 1/6 ,\dots$$

لاحظ أن القيم الممكنة ل X (1، 2، 3، 4، 5، 6) هي متنافية، ولذلك فإن مجموع احتمالاتها يساوي 1.

$$\sum_{x=1}^{6} P(X=x)=1$$

مثال 2. في تجربة إلقاء قطعة نقدية مرتين يمكن أن نعين المتغيرة العشوائية X التي تمثل عدد مرات الحصول على كتابة. في هذه الحالة القيم الممكنة ل X هي 0، 1، 2. لا حظ أنه يمكن تعيين متغيرات عشوائية أخرى انطلاقا منن نفس

[1] في البرنامج الأصلي: 1- مفهوم المتغيرة العشوائية. اخترنا هذا التقسيم لكي يتناسب كل جزء مع الزمن المخصص للمحاضرة.

التجربة، مثلا Y عدد مرات الحصول على صورة، وهي متغيرة تأخذ القيم 0، 1، 2، ثم المتغيرة Z بحث $Z = X - Y$

…

القيم الممكنة ل X هي 0، 2، -2. الاحتمالات الملحقة بقيمها يمكن حسابها كما يلي:

$P(Z = 0) = P(X – Y = 0) = P(X = 0 \text{ et } Y = 0 \text{ ou } X = 1 \text{ et } Y = 1 \text{ ou } X = 2 \text{ et } Y = 2) =>$

$P(Z = 0) = 0 + P(X = 1 \text{ et } Y = 1) + 0 = 2 * 0.5^2 = \mathbf{0.5}$

2 المتغيرة العشوائية المتقطعة

و تسمى أيضا م ع منفصلة، وهي التي تأخذ عددا منتهيا من القيم الممكنة في مجال مغلق.

مثال: داخل المجال المغلق [2, 5] المتغيرة X المعرفة في المثال الأول تأخذ 4 قيم ممكنة.

3 التوزيع الاحتمالي للمتغيرة المتقطعة

هي مجموعة القيم الممكنة مع الاحتمالات المرتبطة بقيم المتغيرة. نرمز للمتغيرة بحرف كبير وللقيم التي تأخذها المتغيرة بحرف صغير. نعبر عن احتمال قيمة معينة كما يلي $P(X = x)$ ونكتب أيضا : $f(x)$. وتسمى الدالة $f(x)$ دالة الكثافة الاحتمالية.

مثال: التوزيع الاحتمالي لم ع للمثال الأول (إلقاء مكعب نرد) يكتب كما يلي:

X	1	2	3	4	5	6	
P(X = x)	1/6	1/6	1/6	1/6	1/6	1/6	1

مثال 2. التوزيع الاحتمالي ل X، عدد مرات الصورة في رميتين لقطعة نقدية:

X	0	1	2	
P(X = x)	1/4	2/4	1/4	1

4 شروط دالة الكثافة للمتغيرة المتقطعة

نعبر عن احتمال قيمة معينة كما يلي $P(X = x)$ ونكتب أيضا : $f(x)$ وتسمى الدالة $f(x)$ دالة الكثافة الاحتمالية.

لكي يمكن اعتبار دالة ما، أيا كانت، دالة كثافة احتمالية يجب أن يتحقق شرطان اثنان:

1) $f(x) \geq 0$

2) $\sum_x f(x) = 1$

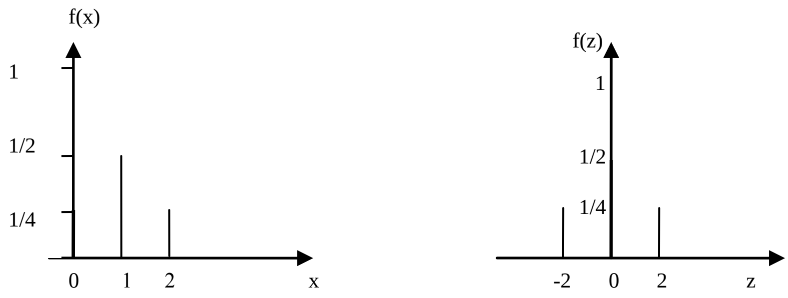
مثال: نأخذ دالة الكثافة ل X نتيجة لإلقاء حجر نرد: $f(1) = f(2) = f(3) = \ldots f(6) = 1/6 \geq 0$,

الشرط الأول محقق، والشرط الثاني أيضا لأن: $\Sigma f(x) = 1/6 + 1/6 + \ldots + 1/6 = 6(1/6) = 1$

5 التمثيل البياني لدالة الكثافة الاحتمالية ل م ع المتقطعة

تمثل المتغيرة العشوائية المتقطعة ليس من خلال منحنى ولكن من خلال أعمدة متوازية على محور X.

مثال: تمثل بيانيا منحنيات دوال الكثافة ل X وZ المعرفة على إلقاء قطعة نقدية مرتين.



رسم 3 التمثيل البياني لدالة الكثافة للمتغيرة العشوائية المتقطعة

6 دالة التوزيع F(x) للمتغيرة العشوائية المتقطعة

تعرف دالة التوزيع – وتسمى أيضا "الدالة التجميعية" – كما يلي:

$$\mathbf{F(x) = P(X \leq x)}$$

ويمكن استنتاج دالة التوزيع من دالة الكثافة الاحتمالية f(x) كما يلي:

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u)$$

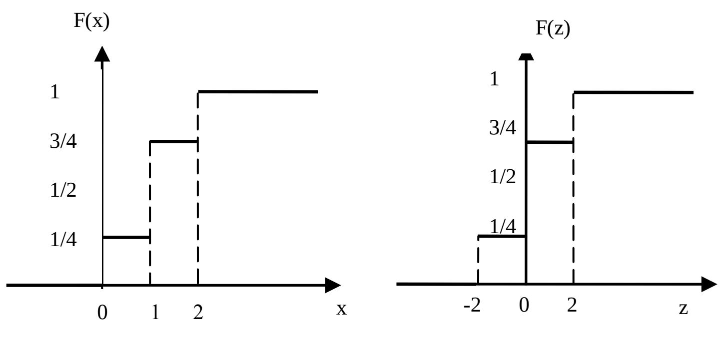
إذا كانت X تأخذ عددا منتهيا من القيم فإن F(x) يمكن تعريفها كما يلي:

$$F(x) = \begin{cases} 0 & , \quad -\infty < x < x_1 \\ f(x_1) & , \quad x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & , \quad x_2 \leq x < x_3 \\ \cdots & \\ f(x_1) + f(x_2) + \ldots + f(x_n), & \quad x_n \leq x < +\infty \end{cases}$$

مثال: أوجد قيم F(x) وF(z) للأمثلة السابقة ومثلهما بيانيا.

X	0	1	2
f(x)	1/4	1/2	1/4
F(x)=P(X≤x)	1/4	3/4	1

Z	-2	0	2
f(x)	1/4	1/2	1/4
F(x)=P(X≤x)	1/4	3/4	1



رسم 4 التمثيل البياني لدالة التوزيع للمتغيرة العشوائية المتقطعة

ملاحظة. تأخذ دالة التوزيع للم ع المتقطعة شكلا سلميا، وهي لا تكون متناقصة في أي مجال، وأكبر قيمة ممكنة لها هي 1.

المبحث 2. مفهوم المتغيرة العشوائية المستمرة وتوزيعها الاحتمالي

تعريف المتغيرة العشوائية المستمرة
التوزيع الاحتمالي للمتغيرة العشوائية المستمرة
خصائص دالة الكثافة الاحتمالية للمتغيرة العشوائية المستمرة
دالة التوزيع للمتغيرة العشوائية المستمرة
قاعدة لايبنيتز

1 تعريف المتغيرة العشوائية المستمرة

هي متغيرة ع تأخذ عددا لا متناهيا من القيم في مجال محدود، أو هي تأخذ أي قيمة داخل هذا المجال. من أجل هذا فأن وحدات قياس المتغيرة الميتمرة تكون مستمرة مثل الزمن، الوزن، المسافة، الحجم، ...

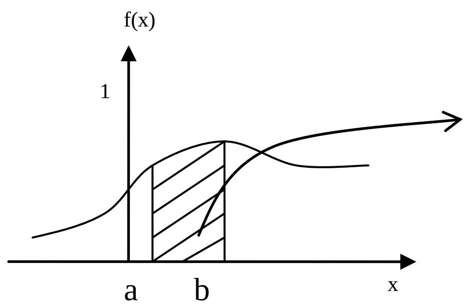
2 التوزيع الاحتمالي المستمر

هو مجموعة القيم التي يمكن أن تأخذها المتغيرة ع المستمرة والاحتمالات الملحقة بها. نسمي توزيعا كهذا دالة الكثافة الاحتمالية أو دالة الاحتمال، وهي ممثلة بمنحنى متصل.

لاحظ أنه بما أن X تأخذ عددا لا متناهيا من القيم داخل أي مجال، فإن احتمال قيمة بعينها هو احتمال يؤول إلى الصفر. $P(X=x) \rightarrow 0$ لذلك فإن دالة الكثافة تستعمل لحساب احتمال مجال. ويكون ذلك بحساب المساحة تحت منحنى f(x) بين حدود المجال.

$$\boxed{P(a < x < b) = \int_a^b f(x)dx}$$

لاحظ أن إشارة التكامل هنا تقابل إشارة المجموع في حالة المتغيرة ع المتقطعة.



رسم5 التمثيل البياني لدالة الكثافة للمتغيرة العشوائية المستمرة

3 خصائص دالة الكثافة الاحتمالية للمتغيرة العشوائية المستمرة

باستبدال إشارة التكامل بإشارة المجموع نجد أن شروط دالة الكثافة الاحتمالية للم ع المستمرة تكتب كما يلي :

$$1)\ f(x) \geq 0$$

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

من البديهي إذا أن منحنى دالة الكثافة لا يمكن أن ينزل أسفل محور الم ع، كما أن المساحة الإجمالية بين المنحنى والمحور الأفقي تساوي الواحد.

هذه الخصائص تفيدنا في حساب احتمالات بعض المجالات من خلال احتمالات مجالات أخرى.

مثال: أوجد قيمة الثابت C التي تحقق الشرطين الأول والثاني لدالة الكثافة الاحتمالية في الدالة التالية:

- ✓ أحسب احتمال أن تكون X تنتمي للمجال من 1 إلى 2.

✓ أحسب احتمال أن تكون X لا تنتمي للمجال من 1 إلى 2.

$$f(x)=\begin{cases} cx^2 & 0<x<3 \\ 0 & \text{sinon} \end{cases}$$

$$\int_{-\infty}^{+\infty} f(x)dx=1 \Rightarrow \int_{-\infty}^{0} 0dx+\int_{0}^{3} Cx^2dx+\int_{3}^{+\infty} 0dx=1 \Rightarrow C\left[\frac{x^3}{3}\right]_0^3=9C=1 \Rightarrow C=1/9$$

لكي تكون x دالة كثافة يجب أن يكون C = 1/9 .

$$P(1<x\le 2)=\int_1^2 f(x)dx=\int_1^2 (1/9)x^2dx=\frac{1}{9}\left[\frac{x^3}{3}\right]_1^2=\frac{1}{9}\left[\frac{8-1}{3}\right]=\frac{7}{27}$$

$$P(1>x>2)=1-P(1<x<2)=1-\frac{7}{27}=\frac{20}{27}$$

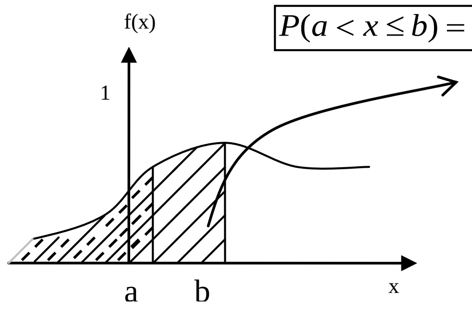
4 دالة التوزيع F(x) للمتغيرة العشوائية المستمرة

تعرف دالة التوزيع للمتغيرة المستمرة كما يلي:

$$\boxed{F(x)=P(X\le x)=\int_{-\infty}^{x} f(u)du}$$

لدالة التوزيع أهمية أكبر بالنسبة للمتغيرة المستمرة. السبب في ذلك أننا نهتم، في حالة المتغيرة المستمرة، باحتمال مجال وليس باحتمال نقطة، ولحساب احتمال مجال من الأيسر التعويض في دالة التوزيع بدلا من حساب التكامل في كل مرة. يتضح ذلك من القاعدة التالية: بفرض a, b نقطتان من مجال تعريف X، بحيث b > a . لحساب احتمال أن تكون X تنتمي إلى المجال]a, b] :

$$\boxed{P(a<x\le b)=P(X\le b)-P(X\le a)=F(b)-F(a)}$$



رسم 6 حساب الاحتمال من خلال دالة التوزيع

مثال:

أوجد دالة التوزيع للمتغيرة المذكورة في المثال السابق.

استخدم دالة التوزيع لحساب الاحتمال: P(1< x <2).

$$*\,x<0: \qquad F(x)=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{0} 0du=0$$

$$*\,0\le x<3: \qquad F(x)=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{0} 0du+\int_{0}^{x}\frac{1}{9}u^2du=\frac{1}{9}\left[\frac{u^3}{3}\right]_0^x=\frac{x^3}{27}$$

$$*\,x\ge 3: \qquad F(x)=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{0} 0du+\int_{0}^{3}\frac{1}{9}u^2du+\int_{3}^{x} 0du=0+\frac{1}{9}\left[\frac{u^3}{3}\right]_0^3+0=\frac{27}{27}=1$$

$$F(x)=\begin{cases} 0 & x<0 \\ x^3/27 & 0\le x<3 \\ 1 & x\ge 3 \end{cases} \qquad P(1<x\le 2)=F(2)-F(1)=\frac{2^3}{27}-\frac{1^3}{27}=\frac{7}{27}$$

5 قاعدة لايبنيز Règle de LEIBNITZ

تفيد هذه القاعدة الرياضية العامة في استنتاج أن مشتقة دالة التوزيع هي دالة الكثافة:

$$\boxed{\frac{d\int_{-\infty}^{x} f(u)du}{dx} = f(x) \Rightarrow \frac{dF(x)}{dx} = f(x)}$$

مثال: أوجد دالة الكثافة للمتغيرة X إذا كانت دالة التوزيع كما يلي:

$$F(x) = \begin{cases} 1-e^{-2x} & x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

$$f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & \text{sinon} \end{cases}$$

$$* x < 0 : f(x) = F'(x) = (0)' = 0$$

$$* x \geq 0 : F(x) = 1-e^{-2x} \Rightarrow f(x) = \left(1-e^{-2x}\right)' = 2e^{-2x}$$

6 خلاصة المبحث الأول و الثاني

يتم تعريف التوزيع الاحتمالي (أو القانون الاحتمالي) لمتغيرة عشوائية من خلال تحديد القيم الممكنة للمتغيرة و الاحتمالات المقابلة لها.

يتم هذا التحديد إما من خلال جدول (جدول التوزيع الاحتمالي) أو دالة، تسمى دالة كثافة الاحتمالية.

لكي نقول عن دالة ما أنها دالة كثافة احتمالية يجب أن تكون موجبة دوما و أن يكون مجموع الاحتمالات مساويا للواحد.

الدالة التجميعية (أو دالة التوزيع) تمثل احتمال مجال من أصغر قيمة للمتغيرة إلى نقطة ما:

$F(x) = P(X \leq x) = \sum_{u \leq x} f(u)$ في حالة م متقطعة و $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$ في حالة م مستمرة.

نظرا لتعريفها تأخذ دالة التوزيع مسارا متزايدا (أو ثابتا على أجزاء من المجال). تبرز أهمية الدالة التجميعية أكثر عندما تكون المتغيرة مستمرة لأننا نهتم حينها باحتمالات مجالات. يمكن استنتاج دالة الكثافة من خلال اشتقاق دالة التوزيع.

الفصل III. التوقع الرياضي والتباين

التوقع الرياضي
التباين والانحراف المعياري
العزوم
الدالة المتجددة للعزوم
نظرية شيبيشيف، نظرية الأعداد الكبيرة

مسألة: أرسلت مؤسسة عروضا إلى 4 عملاء. احتمال تلقي طلبية من العميل الأول هي 0.2، من العميل الثاني 0.3، من العميل الثالث 0.35 و0.4 من العميل الرابع. في انتظار ردود العملاء ما هو العدد المتوقع من الطلبيات؟

في العديد من الحالات لا يكفي حساب احتمال تحقق حدث أو أحداث معينة بل نحتاج للخروج بتوقع معين يلخص الوضعية المطروحة أمامنا. من جهة أخرى قد يصعب المفاضلة بين خيارات متاحة مقيمة بمبالغ معينة بسبب ارتباط كل مبلغ بمخاطرة مختلفة؛ من المعروف أن الاستثمارات الأكثر مردودية هي تلك التي تتضمن أكبر مخاطرة، فكيف يمكن أخذ في الحسبان المخاطرة والمبلغ المتوقع وبطريقة دقيقة وموضوعية؛ إن طريقة التوقع وبقية المفاهيم الأخرى الواردة أعلاه يمكن أن تساعدنا في ذلك.

المبحث 1. التوقع الرياضي Espérance mathématique

تعريف التوقع
توقع دالة

1 تعريف التوقع

يعرف التوقع الرياضي لمتغيرة عشوائية متقطعة كما يلي:

$$\boxed{E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i=1}^{n} x_i f(x_i)}$$

و يعرف التوقع الرياضي لمتغيرة عشوائية مستمرة كما يلي:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

نرمز للتوقع أحيانا ب μ أو μ_X.

مثال: نلقي قطعة نقدية 4 مرات. أحسب العدد المتوقع من المرات التي نحصل فيها على وجه.

المجموع	4	3	2	1	0	**X**عدد مرات وجه
16/16 = 1	$(1/2)^4$	$4\ (1/2)^4$	$6\ (1/2)^4$	$4\ (1/2)^4$	$(1/2)^4$	P(X)
32/16 = 2	$4\ (1/2)^4$	$12\ (1/2)^4$	$12\ (1/2)^4$	$4\ (1/2)^4$	0	XP(X)

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = 32/16 = 2$$

العدد المتوقع هو مرتين من بين 4 رميات.

مثال 2. نلقي قطعة نقدية مرة واحدة. يربح اللاعب 20 دج إذا حصل على الرقم 2، ويربح 40 دج إذا حصل على الرقم 4، و60 دج إذا حصل على الرقم 6، ويخسر 10 دج إذا حصل على الرقم 1، 30 دج إذا حصل على الرقم 3، و50 دج إذا حصل على الرقم 5. تحقق مما إذا كانت العبة متوازنة (هل توقع الربح يساوي توقع الخسارة).

الجواب هو أن اللعبة غير متوازنة لأن توقع الربح أكبر من توقع الخسارة $E(x) = 30/6 = 5 > 0$

نتيجة الرمي	1	2	3	4	5	6	المجموع
X نتيجة المراهنة	-10	20	-30	40	-50	60	-
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6	6/6
X*P(X=x)	-10/6	20/6	-30/6	40/6	-50/6	60/6	(120-90)/6>0

مثال3. أوجد التوقع الرياضي للمتغيرة ذات دالة الكثافة التالية:

$$f(x)=\begin{cases}\frac{1}{2}x & 0<x<2\\ 0 & \text{sinon}\end{cases}$$

$$E(x)=\int_{-\infty}^{\infty} xf(x)dx=\int_{-\infty}^{0} x.0.dx+\int_{0}^{2} x(\frac{1}{2}x)dx+\int_{2}^{\infty} x.0.dx$$

$$E(x)=0+\frac{1}{2}\left[\frac{x^3}{3}\right]_0^2+0=\frac{4}{3}$$

2 توقع دالة

يستخدم توقع دالة عند حساب عدد من المقاييس مثل التباين، العزوم المركزية والعزوم المرتبطة بالأصل.

لتكن X م ع لها دالة كثافة f(x)، و$y = g(x)$ م ع تابعة لها.

$$E(Y)=E(g(x))=\sum_x g(x)f(x)$$

في حالة X م متصلة:

$$E(Y)=E(g(x))=\int_{-\infty}^{+\infty} g(x)f(x)dx$$

مثال1. نلقي قطعة نقدية مرتين. نسمي X عدد مرات الحصول على صورة، و$Y=X^2$. أحسب E(X) وE(Y) .

X	0	1	2	المجموع
P(X = x)	1/4	1/2	1/4	1
X*P(X)	0	1/2	1/2	E(X) =1
X^2	0	1	4	
X^2*P(X)	0	1/2	1	$E(X^2) = 3/2$

مثال2: لتكن X م ع ذات دالة الكثافة التالية، و$Y = g(x) = 3x^2 - 2x$. أحسب E(Y).

$$f(x)=\begin{cases}\frac{1}{2}x & 0<x<2\\ 0 & \text{sinon}\end{cases}$$

$$E(Y)=E(g(x))=\int_{-\infty}^{+\infty} g(x)f(x)dx=\int_{-\infty}^{+\infty}(3x^2-2x)(0)dx+\int_{-\infty}^{+\infty}(3x^2-2x)(x/2)dx+\int_{-\infty}^{+\infty}(3x^2-2x)(0)dx$$

$$E(Y)=0+\int_{0}^{2}(3x^2-2x)(x/2)dx+0=\frac{1}{2}\left[\frac{3x^4}{4}-\frac{2x^3}{3}\right]_0^2=\frac{1}{2}\left[12-\frac{16}{3}\right]=\frac{20}{6}=\frac{10}{3}$$

3 خصائص التوقع الرياضي

$E(C) = C$
$E(CX) = CE(X)$
$E(X+Y) = E(X) + E(Y)$
$E(XY) = E(X)E(Y)$ إذا كانت المتغيرتان مستقلتان.

مثال: تتوقع مؤسسة أن تتلقى كل شهر 3 طلبيات من العميل A و4 من B .

- أحسب العدد المتوقع من الطلبيات المتلقاة من A في السنة.
- أحسب العدد الإجمالي المتوقع من الطلبيات المتلقاة في شهر.

يعين كل عميل من طرفه مندوبا عن كل طلبية لمتابعة إتمامها. كم تتوقع أن يلزم من مقابلة لتعريف مندوبي العميل A بمندوبي B.

$E(12A) = 12E(A) = 12(3) = 36$
$E(A*B) = E(A)*E(B) = 4*3 = 12.$
$E(A + B) = E(A) + E(B) = 3 + 4 = 7$

المبحث 2. التباين والانحراف المعياري Variance et écart type

تعريف التباين
خصائص التباين
المتغيرة المعيارية

1 تعريف التباين

يعرف التباين لمتغيرة عشوائية كما يلي:

$$\boxed{\sigma_X{}^2 = Var(X) = E\left[(X-\mu)^2\right]}$$

و الانحراف المعياري هو جذر التباين. $\sigma = \sqrt{V(X)} = \sqrt{\sigma^2}$.

في حالة المتغيرة العشوائية المتقطعة :

$$V(X) = \sum_x (x-\mu)^2 f(x)$$

في حالة المتغيرة العشوائية مستمرة :

$$V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$

مثال. نلقي قطعة نقدية مرتين. نسمي X عدد مرات الحصول على صورة، أحسب V(X) .

X	0	1	2	المجموع
P(X = x)	1/4	1/2	1/4	1
X*P(X)	0	1/2	1/2	E(X) = μ =1
$(X-\mu)^2$	1	0	1	
$(X-\mu)^2$ * P(X)	1/4	0	1/4	V(X) = 1/2

مثال 2. لتكن X م ع ذات دالة الكثافة التالية؛ أحسب تباين X.

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{sinon} \end{cases}$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx. \qquad \mu = 4/3$$

$$\sigma^2 = \int_{-\infty}^{0}\left(x-\frac{4}{3}\right)^2 * 0dx + \int_{0}^{2}\left(x-\frac{4}{3}\right)^2 \frac{x}{2}dx + \int_{2}^{+\infty}\left(x-\frac{4}{3}\right)^2 0dx = 0 + \frac{1}{2}x\left(x^2 - \frac{8x}{3} + \frac{16}{9}\right)_0^2 + 0$$

$$\sigma^2 = \frac{1}{2}\left(x^3 - \frac{8x^2}{3} + \frac{16x}{9}\right)_0^2 = \frac{1}{2}\left(8 - \frac{32}{3} + \frac{32}{9}\right) = \frac{4}{9}$$

2 خصائص التباين

$V(X) = E(X^2) - E(X)^2$

$V(CX) = C^2V(X)$, $V(C) = 0$

في حالة استقلال المتغيرتان عن بعضهما

$V(X - Y) = V(X) + V(Y)$, $V(X + Y) = V(X) + V(Y)$

مثال: نلقي قطعة نقدية مرتين. نسمي X عدد مرات الحصول على صورة، أحسب $V(X)$ باستخدام الصيغة **$V(X) = E(X^2) - E(X)^2$** .

لتكن المتغيرة $Y = 2X$. أحسب $V(Y)$ ، نلقي حجر نرد ونسمي Z النتيجة المحصل عيها. أحسب تباين المتغيرة W حيث: $W = Z - Y$

X	0	1	2	المجموع
$P(X = x)$	1/4	1/2	1/4	1
$X\ P(X)$	0	1/2	1/2	$E(X) = \mu = 1$
$(X)^2$	0	1	4	
$(X)^2 * P(X)$	0	1/2	1	$E(X)^2 = 3/2$

$V(X) = E(X^2) - E(X)^2 = (3/2) - 1 = 1/2$

$V(Y) = V(2X) = 2^2V(X) = 4\ (1/2) = 2.$

$V(W) = V(Z-2X) = V(Z) + V(2X) = V(Z) + 2^2V(X).$

$V(Z) = E(Z^2) – E(Z)^2 = (1/6)[1^2+2^2+3^2+4^2+5^2+6^2]-(1/6)[1+2+3+4+5+6] = 70/6$

$V(W) = (70/6) + 4\ (1/2) = 82/6 = 13.67$

3 المتغيرة المعيارية Variable centrée réduite

يمكن أن نلحق بأي متغيرة عشوائية X متغيرة معيارية (تسمى أيضا المتغيرة المركزية) ويرمز لها *X . تلحق المتغيرة المعيارية بالمتغيرة الحقيقية من أجل المقارنة لأن المتغيرة المعيارية ليس لها وحدة كالمتر أو الساعة ... وإنما هي تعبر عن كل قيمة x ل X من خلال المسافة بين x والتوقع μ محسوبة لس بالوحدة الأصلية وإنما بالانحرافات المعيارية.

$$\boxed{X^* = \frac{X-\mu}{\sigma}}$$

من خلال خصائص التوقع والتباين نستخرج التوقع والتباين للمتغيرة المعيارية.

$$E(X^*) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}\left[E(X) - E(\mu)\right] = 0$$

$$V(X^*) = E\left[(X^* - E(X^*))^2\right] = E\left[(X^* - 0)^2\right] = E(X^{*2}) = E\left(\frac{(X-\mu)^2}{\sigma^2}\right) = \frac{1}{\sigma^2}E(X-\mu)^2 = \frac{\sigma^2}{\sigma^2} = 1$$

مثال: أحسب *X من أجل متوسط 70، انحراف معياري 5، وX يساوي: 55، 60، 50، 75، 80، 70.

الجواب: القيم هي: -3، -2، -4، 1، 2، 0.

مثال 2. يتدرب عاملان أحمد وعلي من أجل المشاركة في ماراتون عيد العمال 1 ماي. يشترط يوم المسابقة أن يكون وزن المترشح لا يتجاوز المجال $\mu \pm 1.5\sigma$.

إذا كان الوزن المتوسط بالكغ هو $\mu = 70$ والانحراف المعياري هو 5 كغ. هل سيقبل العاملان أحمد وعلي إذا كان وزنهما: 77 كغ، و80 كغ؟

الجواب: مجال القبول هو من 62.5 كغ إلى 77.5 كغ، لذلك فسيرفض علي ويقبل أحمد.

4 خلاصة

التوقع الرياضي و التباين هي أهم المؤشرات المعبرة عن خصائص المتغيرة و يحسبان كما يلي، حسب كون المتغيرة متقطعة أو مستمرة:

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i=1}^{n} x_i f(x_i)$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

لكل من التوقع و التباين 4 خصائص أساسية تتمثل فيما يلي:

$V(C) = 0$	توقع عدد ثابت $E(C) = C$
$V(CX) = C^2V(X)$	$E(CX) = CE(X)$
في حالة استقلال المتغيرتان عن بعضهما $V(X+Y) = V(X) + V(Y)$ $V(X-Y) = V(X) + V(Y)$,	$E(X+Y) = E(X) + E(Y)$
$V(X) = E(X^2) - E(X)^2$	في حالة استقلال المتغيرتان عن بعضهما. $E(XY) = E(X)E(Y)$

من أجل المقارنة بين المتغيرات تستخدم المتغيرة المعيارية التي تسمح بالتعبير عن قيمة x ليس من خلال وحداتها الأصلية (كغ، متر، زمن، ...) وإنما بعدد الانحرافات المعيارية التي تفصل بين القيمة x والتوقع الرياضي.

التوقع الرياضي والتباين للمتغيرة المعيارية $X^* = \dfrac{X-\mu}{\sigma}$ هما على التوالي 0 و 1.

لحساب التوقع الرياضي لدالة ما في X (مثلا التباين، أو X^2) نضرب قيم الدالة في الاحتمالات المقابلة لـX:

$$E(Y) = E(g(x)) = \sum_x g(x) f(x)$$

$$E(Y) = E(g(x)) = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

المبحث 3. العزوم و الدالة المتجددة للعزوم

العزوم
الدالة المتجددة للعزوم

1 Les moments العزوم

كما التباين يعتبر العزم من تطبيقات توقع دالة. نميز بين نوعين من العزوم: العزم المركزي والعزم المرتبط بالأصل. تستخدم العزوم في حساب عدد من المقاييس مثل معامل التماثل α3 (coefficient d'asymétrie) ومعامل التفلطح α4 (Kurtosis ou coefficient d'aplatissement).

(أ) العزم المركزي **Le moment central μr**.

يعرف العزم المركزي من الدرجة r للم ع X كما يلي:

$$\mu_r = E\left((X-\mu)^r\right), \qquad r = 0,\ 1,\ 2,\ ...$$

$$\mu_0 = E\left[(X-\mu)^0\right] = E(1) = 1 \qquad \mu_0 = 1$$

$$\mu_1 = E\left[(X-\mu)^1\right] = E(X) - E(\mu) = 0 \qquad \mu_1 = 0$$

$$\mu_2 = E\left[(X-\mu)^2\right] = V(X) \qquad \mu_2 = \sigma^2$$

يحسب العزم المركزي حسب طبيعة المتغيرة متقطعة أو مستمرة كما يلي:

$$\mu_r = \sum_i (x_i-\mu)^r f(x) \qquad\qquad \mu_r = \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx$$

مثال . أحسب العزوم المركزية من الدرجة 0، 1، 2 و3 للمتغيرة ع ذات دالة الكثافة:

$$f(x) = \begin{cases} x/2 & , \quad 0 < x < 2 \\ 0 & \text{sinon} \end{cases}$$

$$\mu_0 = \int_{-\infty}^{+\infty} (x-\mu)^0 f(x)dx = \int_{-\infty}^{+\infty} 1 f(x)dx = 1$$

$$\mu_1 = \int_{-\infty}^{+\infty} (x-\mu) f(x)dx = \int_{-\infty}^{+\infty} x f(x)dx - \mu\int_{-\infty}^{+\infty} f(x)dx = \mu - \mu(1) = 0$$

$$\mu_2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = \sigma^2 = \frac{4}{9}$$

$$\mu_3 = \int_{-\infty}^{+\infty} (x-\mu)^3 f(x)dx = \int_{-\infty}^{0}\left(x-\frac{4}{3}\right)^3 (0)dx + \int_{0}^{2}\left(x-\frac{4}{3}\right)^3 \frac{x}{2}dx + \int_{2}^{+\infty}\left(x-\frac{4}{3}\right)^3 (0)dx = \frac{16(67)}{27(5)}$$

مثال: أحسب العزوم المركزية من الدرجة 0، 1، 2 و3 للمتغيرة ع المتقطعة التي تمثل عدد مرات الحصول على صورة في رميتين لقطعة نقدية.

(ب) العزم المرتبط بالأصل **Moment autour de la moyenne μ'r**

يعرف العزم المرتبط بالأصل كما يلي:

$$\boxed{\mu_r' = E(X^r), \quad r = 0,1,2,...}$$

$$\mu_0' = E(X^0) = E(1) = 1 \qquad\qquad \mu_0' = 1$$
$$\mu_1' = E(X^1) = E(X) = \mu \qquad\qquad \mu_1' = \mu$$
$$\mu_2' = \mu^2 - \sigma^2 = \mu^2 - \mu_2 \qquad\qquad \mu_2' = \mu^2 - \mu_2$$

مثال: أحسب العزوم المرتبطة بالأصل من الدرجة 0، 1، 2، 3 و4 للمتغيرة ع المتصلة ذات دالة الكثافة:

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{sinon} \end{cases}$$

$$\mu'_0 = 1, \qquad \mu'_1 = \mu = 4/3, \qquad \mu'_2 = \mu^2 - \mu_2 = \left(\frac{4}{3}\right)^2 - \frac{4}{9} = \frac{12}{9},$$
$$\mu'_3 = \int_0^2 x^3 (\frac{x}{2}) dx = \frac{1}{2}\left[\frac{x^5}{5}\right]_0^2 = \frac{16}{5}, \qquad \mu'_4 = \int_0^2 x^4 (\frac{x}{2}) dx = \frac{1}{2}\left[\frac{x^6}{6}\right]_0^2 = \frac{16}{3}$$

مثال2. أحسب العزوم المرتبطة بالأصل من الدرجة 0، 1، 2، 3 و4 للمتغيرة ع المتقطعة التي تمثل عدد مرات الحصول على صورة في رميتين لقطعة نقدية.

ج) العلاقة بين العزم المركزي والعزم المرتبط بالأصل

$$\mu_r = \mu'_r - C_r^1 \cdot \mu'_{r-1} \cdot \mu^1 + \ldots + (-1)^i C_r^i \cdot \mu'_{r-i} \cdot \mu^i + \ldots + (-1)^r \cdot \mu'_0 \cdot \mu^r$$

يمكن أيضا الحصول على العزوم المرتبطة بالأصل من الدرجة r من خلال اشتقاق الدالة المتجددة للعزوم r مرة.

2 الدالة المتجددة للعزوم La fonction génératrice des moments $M_x(t)$

الدالة المتجددة للعزوم هي دالة مرتبطة بمتغيرة (معلمة) t بالإضافة إلى ارتباطها ب X، ود م ع كما يلي:

$$M_x(t) = E(e^{tx})$$

في حالة م ع متقطعة: $M_x(t) = \sum_x e^{tx} f(x)$ و في حالة م ع مستمرة:

$$M_x(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx$$

مثال: أكتب الدالة المتجددة للعزرم من أجل $t \neq 0$ للم ع المعرفة في كما يلي:

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{sinon} \end{cases}$$

$$M_x(t) = \int_{-\infty}^{0} e^{tx}(0)dx + \int_0^2 e^{tx} f(x) dx + \int_2^{+\infty} e^{tx}(0)dx = 0 + \frac{1}{2}\int_0^2 x e^{tx} dx + 0$$

$$M_x(t) = \frac{1}{2}\int_0^2 U dV = \frac{1}{2}\left[[UV]_0^2 - \int_0^2 V dU\right]. \qquad U = x \Rightarrow dU = dx,\ dV = e^{tx}dx \Rightarrow V = \frac{e^{tx}}{t}$$

$$M_x(t) = \frac{1}{2}\left(\left[x\frac{e^{tx}}{t}\right]_0^2 - \int_0^2 \frac{e^{tx}}{t} dx\right) = \frac{1}{2}\left(\left[2\frac{e^{2t}}{t}\right] - \frac{1}{t}\left[\frac{e^{2t}}{t}\right]\right) = \frac{e^{2t}}{t} - \frac{e^{2t}}{2t^2}$$

تستخدم الدالة المتجددة للعزوم لحساب العزم المركزي من درجة r:

$$\boxed{\mu'_r = \frac{d^r M_x(t)}{dt^r} \ \ avec\ t = 0}$$

كما تستخدم الد م ع لإثبات تساوي توزيعين احتماليين، مثلا عند تحقق شروط معينة، ونحتاج ذلك خاصة عند دراسة التقارب بين التوزيعات الاحتمالية. النظرية التي نستعملها في ذلك هي كالآتي:

لتكن م ع X وY لهما الد م ع $M_x(t)$ و $M_y(t)$؛ نقول أن م ع X وY لهما نفس التوزيع الاحتمالي إذا!:

$$\mathbf{M_x(t) = M_y(t)}$$

كما تستخدم الد م ع لإثبات إستقلال توزيعين احتماليين. النظرية التي نستعملها في ذلك هي كالآتي:

إذا X وY م ع مستقلتان، لهما الد م ع $M_x(t)$ و $M_y(t)$؛ فإن: $\mathbf{M_{x+y}(t) = M_x(t) . M_y(t)}$

3 خلاصة

العزم و الدالة المتجدد للعزوم هي عبارة عن توقعات دوال (أنظر المبحث السابق). يدخل العزم في حساب بعض المؤشرات مثل التباين و التوقع الرياضي، معامل التفلطح و معامل التماثل.

يعرف العزم المركزي من الدرجة r للم ع X كما يلي: $r = 0, 1, 2, \ldots$ $\mu_r = E\left((X-\mu)^r\right)$,

يعرف العزم المرتبط بالأصل كما يلي: $\mu'_r = E(X^r), \quad r = 0, 1, 2, \ldots$

تعرف الدالة المتجددة للعزوم كما يلي: $M_x(t) = E(e^{tx})$

تستخدم الدالة المتجددة للعزوم لإثبات التقارب بين توزيعات احتمالية و ذلك من خلال نظريتين أساسيتين.

- نقول أن م ع X وY لهما نفس التوزيع الاحتمالي إذا!: $M_x(t) = M_y(t)$
- إذا X وY م ع مستقلتان فإن: $M_{x+y}(t) = M_x(t) . M_y(t)$

المبحث 4. نظرية شيبيشيف ونظرية الأعداد الكبيرة

1 متراجحة شيبيشيف Inégalité de Bienaymé CHEBYCHEV

هي نظرية تخص م ع المتقطعة والمستمرة على السواء، التي يكون لها متوسط وتباين محدود. تستخدم هذه النظرية في قياس التشتت حول التوقع μ، وذلك عن طريق احتمال (أو نسبة) المفردات التي المسافة (الفرق) بينها وبين μ تزيد عن مقدار ما: $P(-\varepsilon \geq (x - \mu) \geq \varepsilon)$ أو بعبارة أكثر اختصارا: $P(|x - \mu| \geq \varepsilon)$. حسب نظرية شيبيشيف فإن هذه النسبة لا تزيد عن σ^2/ε^2، وذلك مهما كانت طبيعة التوزيع. ونعبر عن هذه النظرية كما يلي:

إذا كانت X م ع متصلة أو متقطعة، لها متوسط μ وتباين محدود σ^2، فإنه مهما يكن ε عدد موجب تماما:

$$\boxed{P(|x-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}}$$

و من أجل صياغة أكثر دلالة، نضع: ε = k σ فنجد : $\boxed{P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}}$

- لاحظ أنه، انطلاقا من نفس النتيجة، لدينا أيضا:

$$P(-\varepsilon < (x-\mu) < \varepsilon) = P(|x-\mu| < \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2}$$

سؤال: هل هذه العبارة صحيحة: إذا كان أقل من 10 % من الفائزين في امتحان الباكالوريا يحصلون على تقدير أكثر من حسن، فهذا يعني أن أكثر من 90% من الفائزين في الباكالوريا يحصلون على تقدير أقل من حسن.

الجواب نعم.

مثال: لتكن X م ع تتبع توزيعا أيا كان؛ لها متوسط μ وتباين محدود σ^2، وε عدد موجب تماما؛

1. أحسب الحد الأقصى للاحتمال $P(-\varepsilon \geq (x - \mu) \geq \varepsilon)$ من أجل $\varepsilon = 2\sigma$، $\varepsilon = 3\sigma$، $\varepsilon = 4\sigma$.
2. أحسب الحد الأدنى للاحتمال $P(-\varepsilon < (x - \mu) < \varepsilon)$ من أجل $\varepsilon = 2\sigma$، $\varepsilon = 3\sigma$، $\varepsilon = 4\sigma$.

$$P(-\varepsilon \geq (x-\mu) \geq \varepsilon) = P(|x-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

$$\varepsilon = 2\sigma \Rightarrow P(-\varepsilon \geq (x-\mu) \geq \varepsilon) = P(|x-\mu| \geq 2\sigma) \leq \frac{1}{4}, \qquad \varepsilon = 3\sigma \Rightarrow P(|x-\mu| \geq 3\sigma) \leq \frac{1}{9},$$

$$\varepsilon = 4\sigma \Rightarrow P(-\varepsilon \geq (x-\mu) \geq \varepsilon) = P(|x-\mu| \geq 4\sigma) \leq \frac{1}{16},$$

$$P(-\varepsilon < (x-\mu) < \varepsilon) = P(|x-\mu| < \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2}$$

$$\varepsilon = 2\sigma \Rightarrow P(-\varepsilon < (x-\mu) < \varepsilon) = P(|x-\mu| < 2\sigma) > 1 - \frac{1}{4} = 3/4, \qquad \varepsilon = 3\sigma \Rightarrow P(|x-\mu| < 3\sigma) > 1 - \frac{1}{9} = 8/9,$$

$$\varepsilon = 4\sigma \Rightarrow P(-\varepsilon < (x-\mu) < \varepsilon) = P(|x-\mu| < 4\sigma) > 1 - \frac{1}{16} = 15/16,$$

مثال 2. يتم أخذ أوزان عمال مركب الحجار من أجل ترشيحهم للمشاركة في ماراتون عيد العمال 1 ماي.

1. كيف يمكن تحديد مجال القبول بحيث يتم ترشيح على الأقل 75% من العمال، رفض ترشيح أقل من 25% من العمال؟
2. حدد قيم المجال إذا كان الوزن المتوسط الافتراضي هو 70 كغ، والانحراف المعياري 5كغ.

$$P(-\varepsilon < (x-\mu) < \varepsilon) = P(|x-\mu| < \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2}$$

$$\text{Lorsque } \varepsilon = 2\sigma. \text{ nous savons déjà que} P(-\varepsilon < (x-\mu) < \varepsilon) = P(|x-\mu| < 2\sigma) > 1 - \frac{1}{4} = 75\%$$

$$P(\varepsilon \geq (x-\mu) \geq \varepsilon) = P(|x-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

$$\text{Lorsque } \varepsilon = 2\sigma. \text{ nous savons déjà que } P(|x-\mu| \geq 2\sigma) \leq \frac{1}{4} = 25\%$$

1. إن أكثر من 75 % العمال لهم أوزان لا تبتعد بأكثر من 2 σ عن المتوسط، وهذا يمكن التعبير عنه بطريقة أخرى بالقول: إن أقل أو يساوي من 25 بالمئة من العمال لهم أوزان أبعد من الوزن المتوسط بأكثر من 2σ.
   إذا يمكن اتخاذ مجال قبول μ±2σ لتحقيق الهدف المسطر.
2. المجال الذي يحقق الهدف حسب القيم المعطاة للمتوسط والانحراف المعياري هو من 60 إلى 80 كغ.

مثال 3. متوسط مدة التمدرس في مجتمع معين هي 8 سنوات، والانحراف المعياري $\sigma = 1$.

1. أحسب أدنى احتمال لمدة تمدرس بين 6 و10 سنوات لفرد مختار عشوائيا من هذا المجتمع.
2. أحسب أقصى احتمال لمدة تمدرس لا تزيد عن 6 سنوات أو لا تقل عن 10 سنوات.

$$P(|x-\mu| < k\sigma) > 1-\frac{1}{k^2}. \qquad \mu = 8, \sigma = 1 \Rightarrow 10 = \mu + 2\sigma,\ 6 = \mu - 2\sigma. Pour\ k = 2:$$

$$P(10 > x > 6) = P(|x-\mu| < 2\sigma) > 1-\frac{1}{2^2} = 3/4$$

أدنى احتمال لمدة تمدرس بين 6 و10 سنوات لفرد مختار عشوائيا من هذا المجتمع **هي** 0.75.

$$P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}. \qquad 6 = \mu - 2\sigma.\ soit\, k = 2:$$

$$P(|x-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = 1/4$$

أدنى احتمال لمدة تمدرس لا تزيد عن 6 سنوات أو لا تقل عن 10 هو 0.25.

2 نظرية الأعداد الكبيرة Théorème des grands nombres

تعتبر نظرية الأعداد الكبيرة من نتائج نظرية شيبيشيف ويستفاد منها بشكل خاص في نظرية المعاينة. تصاغ هذه النظرية بالشكل الآتي: لتكن المتغيرات X_1، X_2، . . . . متغيرات عشوائية مستقلة لها نفس التوزيع الاحتمالي ولكل منها نفس المتوسط μ والتباين σ^2 إذا كانت

$$S_n = X_1 + X_2 + \ldots + X_n \qquad (n = 1, 2, \ldots),$$

$$\lim_{n \longrightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

بما أن $E(S_n/n) = \mu$ فإن مضمون هذه النظرية هو أن احتمال أن تبتعد المتغيرة S_n/n عن قيمتها المتوقعة بأكثر من ε هو 0 عندما $n \to \infty$. تسمى هذه الصياغة أيضا بقانون الأعداد الكبيرة الضعيف. حيث أن قانون الأعداد الكبيرة القوي هو:

$$P[\lim_{n\to\infty} S_n/n = \mu] = 1$$

3 خلاصة

نظرية شيبيشاف ونظرية الأعداد الكبيرة من النظريات التي تقيس تشتت المتغيرة و هي من تطبيقات توقع دالة:

- إذا كانت X م ع متصلة أو متقطعة، لها متوسط μ وتباين محدود σ^2، فإنه مهما يكن ε عدد موجب تماما:

$$P(|x-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

- لتكن المتغيرات X_1، X_2، . . . . متغيرات عشوائية مستقلة لها نفس التوزيع الاحتمالي ولكل منها نفس المتوسط μ والتباين σ^2 إذا كانت

$$S_n = X_1 + X_2 + \ldots + X_n \qquad (n = 1, 2, \ldots),$$

$$\lim_{n \longrightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

الفصل IV. التوزيعات الاحتمالية الأكثر استخداما

التوزيعات الاحتمالية المتقطعة
التوزيعات الاحتمالية المستمرة

المبحث 1. التوزيعات لاحتمالية المتقطعة الأكثر استخداما

التوزيع الهندسي الزائد، التوزيع الهندسي الزائد المتعدد، توزيع برنولي، التوزيع الثنائي، التوزيع الثنائي السالب (باسكال)، التوزيع الهندسي، التوزيع المتعدد، توزيع بواسون.

بعد أن عرفنا مفهوم المتغيرة العشوائية والتوزيع الاحتمالي يمكننا الآن أن ندرس عدد من التوزيعات الاحتمالية الشهيرة. تستخدم هذه التوزيعات في حل العديد من المسائل في مجال التسيير الصناعي والتجاري وفي الإدارة. ومن أكثر هذه التوزيعات شيوعا: التوزيع الثنائي وتوزيع بواسون. في نهاية المحاضرة يفترض أن يكون الطالب قادرا على استذكار القوانين المدروسة وخصائصها الأساسية، ومن خلال التطبيقات يفترض أن يتمكن من معرفة متى وكيف يمكن استخدام كل قانون.

1 التوزيع الهندسي الزائد: Distribution hyper géométrique

(أ) استنتاج صيغة قانون التوزيع الهندسي الزائد:

مثال 1. صندوق به 6 كريات منها 4 بيضاء و2 حمراء. نسحب بدون إرجاع 3كريات. احسب احتمال الحصول على كرتين بيضاوين، 3كريات بيضاء، كرية واحدة بيضاء، ولا كرية بيضاء.

نفترض أننا نسحب من صندوق كريات **بدون إرجاع** عددها n، إذا كان الصندوق يحتوي على N كرية منها b بيضاء وr حمراء $(N = b + r)$ فإن احتمال الحصول على عدد معين $x \leq b$ من الكريات البيضاء يمكن أن نحصل عليه من خلال القانون الكلاسيكي للاحتمالات (ع الحالات الملائمة / ع الحالات الممكنة) وذلك باستخدام التوفيقات:

$$\boxed{P(X = x) = \frac{C_b^x \cdot C_r^{n-x}}{C_N^n}}$$

تسمى هذه الصيغة: قانون التوزيع الهندسي الزائد ونكتب **$X \sim H(N, b, p)$** حيث:

$p = b/N$ و $q = r/N = 1-p$

يمكن الآن الإجابة على أسئلة المثال كما يلي:

$P(X=2) = C_4^2 . C_2^1 / C_6^3 = 12/20$, $P(x = 3) = C_4^3 . C_2^0 / C_6^3 = 1/5$, …

(ب) خصائص التوزيع الهندسي الزائد

$$\mu = np, \qquad \sigma^2 = npq\left(\frac{N-n}{N-1}\right)$$

ملاحظة: للتوزيع الهندسي الزائد علاقة بالتوزيع الثنائي سنذكرها عندما نتطرق لهذا الأخير.

2 التوزيع الهندسي الزائد المتعدد: Distribution Multi-hypergéométrique

(أ) استنتاج صيغة قانون التوزيع الهندسي الزائد المتعدد

يمكن بسهولة تعميم القانون السابق على حالة وجود أكثر من صنفين (k صنف)، حيث من كل صنف لدينا Ni كرية، ($\Sigma Ni = N$) ، ولحساب احتمال نتيجة معينة؛ مثلا 2 كريات بيضاء ($X_1 = 2$) ، 5حمراء، 1 زرقاء، . . . يمكن حساب عدد الحالات الملائمة والممكنة من خلال التوفيقات كما يلي:

$$\boxed{P(X_1 = x_1, X_2 = x_2, \dots X_k = x_k) = \frac{C_{N_1}^{x_1} C_{N_2}^{x_2} \dots C_{N_k}^{x_k}}{C_N^n} \qquad \sum_1^k N_i = N, \sum_1^k x_i = n}$$

(ب) خصائص التوزيع الهندسي الزائد المتعدد

$$E(X_i) = n\frac{N_i}{N} = np_i$$

ملاحظة : للتوزيع الهندسي الزائد علاقة بالتوزيع الثنائي سنذكرها عندما نتطرق لهذا الأخير.

3 توزيع برنولي[1] Distribution de Bernoulli

(أ) استنتاج صيغة قانون برنولي

نقول عن تجربة أنها "برنولية" إذا كانت تحتمل نتيجتين (حدثين) متنافيتين A و'A. نسمي A نجاح و'A فشل.

نعتبر المتغيرة ع X التي تمثل **عدد مرات النجاح**، تأخذ X القيمة 1 عند تحقق الحدث A و0 في الحالة المعاكسة.

نرمز عادة ب p "احتمال النجاح" لاحتمال تحقق الحدث A و $q = 1 - p$ احتمال الحدث المعاكس (الفشل). يعين توزيع برنولي كما يلي :

$$\boxed{P(X = 1) = p, \qquad P(X = 0) = q, \qquad X = 0, 1.}$$

ونكتب **$X \sim B(1, p)$**

(ب) خصائص توزيع برنولي

$$E(X) = \Sigma x_i p_i = 1.p + 0.q = p \quad \Rightarrow \quad E(X) = p.$$

$$V(X) = E(X^2) - E(X)^2 = (1^2.p + 0^2.q) - p^2 = p - p^2 = p(1-p) = pq \Rightarrow \quad \mathbf{V(X) = qp.}$$

الدالة المتجددة للعزوم. $M(t) = E(e^{xt}) = e^{0t}q + e^{1t}p \Rightarrow \quad \mathbf{M(t) = q + pe^t}$

معامل التماثل

$$\mu_3 = \sum (x-\mu)^3 p(x) = (0-p)^3 q + (1-p)^3 p = pq^3 - qp^3 = qp(q^2 - p^2)$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{qp(q^2 - p^2)}{qp\sqrt{qp}} \qquad = \frac{q^2 - p^2}{\sqrt{qp}}$$

[1] بإسم Jacques Bernoulli الذي درس هذا التوزيع في أواخر القرن 17.

4 التوزيع الثنائي Distribution binomiale

(أ) استنتاج صيغة قانون التوزيع الثنائي:

إذا كررنا تجربة برنولي n مرة فإن X (عدد مرات النجاح) تأخذ القيم: X = 0, 1, 2, 3, . . . n

لنفترض التجربة البرنولية رمي قطعة نقدية مكررة عدد n من المرات، وX عدد مرات الحصول على صورة (F):

حالة : n = 2 X = 0, 1, 2.

$P(X = 0) = q*q = q^2$, $P(X=1) = P(FP) + P(PF) = p*q + q*p = \mathbf{2p^1q^1}$

حالة : n = 3 X = 0, 1, 2, 3.

$P(X=3) = P(FFF) = p*p*p = p^3$, $P(X=2) = P(FFP \text{ ou } PFF \text{ ou } FPF) = \mathbf{3p^2q^1}$

حالة : n = 4 X = 0, 1, 2, 3, 4.

$P(X=3) = P(FFFP \text{ ou } PFFF \text{ ou } FPFF \text{ ou } FFPF) = \mathbf{4\, p^3q^1}$

في النتيجة الأخيرة نلاحظ العدد 3 هو x ، العدد 1 هو n-x ، والعدد 4 هو عدد الطرق الملائمة للحصول على ثلاث نجاحات من بين (n=4) تجارب، ويمكن حسابه كما يلي:

$$C_n^x = \frac{n!}{x!(n-x)!}$$

وبالتالي فاحتمال عدد ما x من النجاحات من بين n تجربة برنولية يحسب كما يلي:

$$\boxed{P(X = x) = C_n^x p^x q^{n-x}, \qquad x = 0,1,2,3,\ldots,n \qquad , \qquad n = 1,2,3,\ldots}$$

حيث x عدد مرات النجاح، p احتمال النجاح في التجربة الواحدة (يبقى ثابت عند تكرار التجربة)، q = 1-p احتمال الفشل وn عدد التجارب. و هو تعريف " قانون التوزيع الثنائي" ويكتب قانون التوزيع الاحتمالي أيضا كما يلي:

$$P(X = x) = C_n^x p^x (1-p)^{n-x}$$

أو **X ~ B(n, p)**.

(ب) شروط استخدام التوزيع الثنائي

- تجربة برنولية مكررة عدد محدد من المرات
- احتمال النجاح في التجربة ثابت (التجارب مستقلة)

مثال : أحسب عند رمي قطعة نقدية متوازنة 4 مرات احتمال الحصول على:

ولا مرة صورة، مرة واحدة، مرتين، 3 مرات، 4 مرات.

$P(X = x) = C^x_n\ p^x\ q^{n-x} \Rightarrow P(X = 0) = C^0_4\ 0.5^0\ 0.5^4 = 1/16$

$P(X = 1) = C^1_4\ 0.5^1\ 0.5^3$ $\qquad P(X = 2) = C^2_4\ 0.5^2\ 0.5^2$

مثال 2: نسحب بالإرجاع 3 كريات من صندوق يحتوي 5 كريات منها 3 حمراء.

أحسب احتمال الحصول على كريتين حمراء.

$P(X=2) = C^2_3\ (3/5)^2\ (2/5)^1$

(ج) خصائص التوزيع الثنائي

التوقع والتباين: يمكن اعتبار X مجموع متغيرات مستقلة برنولية $X = X_1 + X_2 + \dots X_i + \dots + X_n$ لها نفس المعلم p وبالتالي نفس التوقع $(E(Xi) = p)$ أيضا. إذا باستخدام خصائص التوقع والتباين نجد:

$$E(X) = E(X_1 + X_2 + \dots X_i + \dots + X_n) = \Sigma E(Xi) = \Sigma pi = n\,p \Rightarrow \mathbf{E(X) = np}$$

$$V(X) = V(X_1 + X_2 + \dots X_i + \dots + X_n),$$

Xi مستقلة إذن $V(X) = \Sigma V(Xi) = \Sigma pq \Rightarrow \mathbf{V(X) = npq}$

مثال: أحسب التوقع والتباين للمثال السابق :

الدالة المتجددة للعزوم: باعتبار X مجموع متغيرات برنولية مستقلة $X = X_1 + X_2 + \dots X_i + \dots + X_n$ لها نفس المعلم p ونفس الدالة المتجددة للعزوم: $\mathbf{M_X(t) = [q + pe^t]}$ وباستخدام النظرية السابقة بخصوص الدالة م للعزوم: "من أجل X_1 وX_2 م ع مستقلة لها الدالة م للعزوم $M_{x1}(t)$ و$M_{x2}(t)$ فإن: $M_{x1+x2}(t) = M_{x1}(t).\,M_{x2}(t)$" ؛ نستنتج:

$$M_X(t) = M_{x=x1+x2+\dots xn}(t) = M_{x1}(t)\,.\,M_{x2}(t)\dots M_{xn}(t)$$

$$M_X(t) = E(e^{x1t})\,.\,E(e^{x2t})\dots E(e^{xnt}) \Rightarrow \mathbf{M_X(t) = [q + pe^t]^n}$$

معامل التماثل

$$\alpha_3 = \frac{\mu_3}{\sigma^3}. \qquad \mu_3 = \sum (x-\mu)^3 p(x) = \sum (x-np)^3 p(x) = \dots = npq(q-p)$$

$$\Rightarrow \alpha_3 = \frac{npq[(1-p)-p]}{npq\sqrt{npq}} = \frac{1-2p}{\sigma} \quad ou\,encore: \quad \alpha_3 = \frac{q-p}{\sqrt{npq}}$$

يكون منحنى التوزيع الثنائي متماثلا عندما $\alpha_3 = 0 \Rightarrow 2\,p = 1 \Rightarrow p = ½$

معامل التفلطح

$$\alpha_4 = 3 + \frac{1-6pq}{\sqrt{npq}}$$

يكون منحنى التوزيع معتدلا عندما $\alpha_4 = 3 \Rightarrow qp = 1/6$

(د) قاعدة تقارب: العلاقة بين التوزيع الهندسي الزائد والتوزيع الثنائي

في حالة N كبير جدا (يؤول إلى ∞) فإن $(N-n) / (N-1)$ تؤول إلى 1 (n محدود). ومن جهة أخرى يعطي التوزيع الثنائي نتائج قريبة من التوزيع الهندسي الزائد ويصبح السحب بدون إرجاع مطابقا تقريبا للسحب بالإرجاع.

5 التوزيع الثنائي السالب (باسكال) Distribution binomiale négative

(أ) استنتاج صيغة قانون التوزيع الثنائي السالب:

مثال: نرمي قطعة نقود إلى غاية الحصول على 3 مرات صورة (متتالية أو لا). أحسب احتمال أن نحصل على ذلك بعد 5 رميات، 4 رميات، 3 رميات، توقع عدد الرميات اللازمة وأحسب التباين.

من جديد ليكن لدينا تجربة برنولية (نتيجتين نجاح وفشل) مكررة، لكن هذه المرة **إلى غاية الحصول على عدد معين (r) من النجاحات.** X في هذه الحالة هي **عدد مرات تكرار التجربة** إلى غاية الحصول على r نجاح.

كيف يحسب الاحتمال ؟ نعلم أن تحقق النجاح r مرة احتماله p^r واحتمال الفشل x-r مرة يساوي q^{x-r} . إذا الاحتمال المطلوب يتضمن جداء هذين الاحتمالين $p^r q^{x-r}$. لكن هناك عددا من الطرق الملائمة لتحقيق r نجاح من بين X تجربة مع العلم أن آخر تجربة هي نجاح. هذا العدد يساوي إذا عدد الطرق الملائمة لاختيار r-1 نجاح من بينx-1 تجربة C^{r-1}_{x-1} (التجربة الأخيرة معلومة النتيجة).

$$\boxed{P(X=x)=C^{r-1}_{x-1}p^r q^{x-r}, \qquad X=r,r+1,r+2,\ldots+\infty, \qquad r=1,2,3,\ldots,+\infty}$$

يسمى هذا التوزيع توزيع باسكال أو الثنائي السالب ونكتب: **X~B (N, r, p)**

يمكن إذا الإجابة على أسئلة المثال السابق بما يلي:

$$P(X=5)=C^{3-1}_{5-1}\ p^3 q^{5-3}=C^2_4\ (½)^3(½)^2=6\ (1/8)\ (1/4)=\mathbf{9/32}$$

$$\mu=r/p=3/(1/2)=\mathbf{6}\ ,\quad \sigma^2=rq/p^2=3\ (1/2)\ /\ (1/2)^2=12/2=\mathbf{6}$$

(ب) خصائص التوزيع الثنائي السالب

$$\mu=\frac{r}{p}, \qquad \sigma^2=\frac{rq}{p^2}, \qquad M(t)=p\frac{e^t}{\left(1-qe^t\right)^r}$$

$$\alpha_3=\frac{q+1}{\sqrt{q}}\ ,\quad \alpha_4=3+\frac{(q+2)^2+3(nq-1)}{nq}$$

6 التوزيع الهندسي Distribution géométrique

أ) استنتاج صيغة قانون التوزيع الهندسي

نرمي قطعة نقدية إلى أن نحصل على صورة. احتمال أن يتطلب ذلك 4 رميات هو: P(X= 4) = P(PPPF)

نعود من جديد إلى التجربة البرنولية وهذه المرة نكرر التجربة إلى غاية الحصول على النتيجة أو الحدث المطلوب (نجاح مرة واحدة). المتغيرة العشوائية X التي تمثل عدد مرات تكرار التجربة (بما فيها المرة التي حصل فيها النجاح) تتبع التوزيع الهندسي.

إذا رمزنا لاحتمال النجاح ب p ولاحتمال الفشل ب q فإن الاحتمال يمكن كتابته كما يلي: $P(X=4)=q^3p$

وبصفة عامة فإن احتمال أي قيمة لـX يعبر عنه كما يلي :

$$\boxed{P(X=x)=q^{x-1}p \qquad , \qquad X=1,2,3,\ldots}$$

(ب) خصائص التوزيع الهندسي

$$\mu=\frac{1}{p}, \qquad \sigma^2=\frac{q}{p^2}, \qquad M(t)=p\frac{e^t}{\left(1-qe^t\right)} \qquad \alpha_3=\frac{q+1}{\sqrt{q}}\ , \qquad \alpha_4=12+\frac{p^2}{q}$$

ملاحظة: التوزيع الهندسي ما هو إلا حالة خاصة من توزيع باسكال حيث r = 1

7 التوزيع المتعدد Distribution multinomiale

(أ) استنتاج صيغة قانون التوزيع الثنائي المتعدد

مثال. نرمي قطعة نرد 4 مرات. أوجد احتمال الحصول على مرتين الرقم 6 ومرتين الرقم 1.

التوزيع المتعدد هو تعميم للتوزيع الثنائي، فبينما الأول يستعمل في حالة تجربة تقبل نتيجتين فقط، يستعمل التوزيع المتعدد للحالة العامة حيث يكون للتجربة عدد k من النتائج الممكنة. مع استقلالية التجارب عن بعضها. نرمز لهذه النتائج ب $A_1, A_2, \ldots A_k$ ولاحتمالاتها ب $p_1, p_2, p_3, \ldots p_k$. بما إن الأحداث (النتائج) Ai متنافية فإن:

$$p_1+p_2+p_3+\ldots+p_k=1$$

إذا كررنا هذه التجربة **متعددة** النتائج عدد n من المرات فسيكون لدينا لكل حدث (نتيجة) متغيرة عشوائية تمثل عدد مرات وقوعه. نرمز لهذه المتغيرات ب $X_1, X_2, \ldots X_k$ حيث $X_1+X_2+\ldots+X_k=n$.

يحسب احتمال الحدث المركب: $X_1=x_1, X_2=x_2, \ldots, X_k=x_k$ كما يلي :

$$\boxed{P(X_1=x_1, X_2=x_2, \ldots, X_k=x_k)=\frac{n!}{x_1!x_2!\ldots x_k!}p_1^{x_1}p_2^{x_2}\ldots p_k^{x_k}}$$

(ب) خصائص التوزيع المتعدد

$$E(X_1)=np_1, E(X_2)=np_2, \ldots, \quad \mathbf{E(X_k)=np_k}$$
$$V(X_1)=np_1q_1, V(X_2)=np_2q_2, \ldots \quad \mathbf{V(X_k)=np_kq_k}$$

(ج) العلاقة مع التوزيع الهندسي الزائد المتعدد

في التوزيع الهندسي الزائد المتعدد، عندما $N\rightarrow\infty, Ni\rightarrow\infty, Ni/N \rightarrow pi$ ؛ يستخدم التوزيع المتعدد لحساب الاحتمالات.

مثال: إذا رمينا قطعة نرد 42 مرة، أحسب احتمال أن يظهر كل رقم عدد من المرات يتناسب مع الرقم ذاته (الرقم 1 يظهر مرتين، الرقم 2 يظهر 4 مرات، الرقم 3 يظهر 6 مرات وهكذا).

$$P(X_1=2,\ X_2=4, \ldots X_6=12)=\frac{42!}{2!\ 4!\ 6!\ \ldots\ 12!}(1/6)^2(1/6)^4\ \ldots(1/6)^{12}$$

مثال3. نسحب من صندوق به 5 كريات مرقمة من 1 إلى 5 سبع مرات على التوالي كرية ثم نرجعها إلى الصندوق. أوجد احتمال: 3 كريات ذات رقم 1، كريتين ذات رقم 2 وكريتين ذات رقم 4.

8 توزيع بواسون Distribution de Poisson [1]

(أ) استنتاج صيغة قانون توزيع بواسن

لتكن لدينا تجربة برنولية مكررة عدد كبير جدا أو لانهائي من المرات. مبدئيا المتغيرة X التي تمثل عدد النجاحات تتبع التوزيع الثنائي، لكن قد يصعب حساب الاحتمال باستعمال صيغة هذا التوزيع عندما تكون n كبيرة. مثلا إحتمال 20 نجاح إذا كانت n = 100 هو: $P(20) = C_{100}^{20} \cdot 0.001^{20} \cdot 0.999^{80}$.

عندما تتكرر التجربة باستمرار؛ يصبح عدد مرات تكرار التجربة مقاسا بالزمن، ويكون احتمال تحقق الحدث في لحظة زمن صغيرا جدا. نحتاج في هذه الحالة إلى إيجاد صيغة عامة تعادل صيغة التوزيع الثنائي عندما n يؤول إلى ∞ .

نضع λ ثابت بحيث $p = \lambda/n$:

$$p(x) = C_n^x \cdot p^x \cdot q^{n-x} = \frac{n!}{(n-x)!\,x!} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$p(x) = \frac{n(n-1)(n-2).....(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$p(x) = \frac{\frac{n}{n}\,\frac{(n-1)}{n}\,\frac{(n-2)}{n}.....\frac{(n-x+1)}{n}}{x!} \lambda^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right).....\left(1-\frac{x-1}{n}\right)}{x!} \lambda^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$n \to \infty \Rightarrow \frac{1}{n} = \frac{2}{n} = ... = 0 \Rightarrow p(x) = \frac{1}{x!}\lambda^x \left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}$$

لكن : $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$ وبما أن $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = (1-0)^{-x} = 1$ فإن:

$$\boxed{p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0,1,2,...}$$

و هو احتمال x نجاح في وحدة زمن واحدة حسب توزيع بواسون حيث $\lambda > 0$. ونكتب **X~P(λ)**

ملاحظة: $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e = 2.71828...$

(ب) خصائص توزيع بواسون

$$E(X) = V(X) = \lambda \quad , \quad M(t) = \exp\left[\lambda(e^t - 1)\right] \ , \quad \alpha_3 = \frac{1}{\sqrt{\lambda}}, \quad \alpha_4 = 3 + \frac{1}{\lambda}$$

---

[1] باسم سيميون دونيز بواسون Siméon-Denis Poisson (1781-1840) الفيزيائي و الرياضي الفرنسي الذي استخدم هذا لقانون سنة 1837 في كتابه بحث في احتمال الأحكام في مجال الجريمة و في المجال المدني (Recherche sur la probabilité des jugements en matière criminelle et en matière civile) حيث أدخل كنهاية لقانون باسكال والقانون الثنائي. إلا أن أول استعمال له للقانون الذي يحمل اسمه يعود إلى 1830. تجدر الإشارة إلى أن بواسون صاحب الفضل في نظرية مهمة أخرى هي نظرية الأعداد الكبيرة التي تنسب لشيبيشيف. أنظر ج ج دراوزبيك [1997].

(ج) حساب احتمال عدد من الأحداث في t وحدة زمن.

من أجل عدد أو مقدار t من وحدات الزمن نعوض λ ب λt فنجد:

$$P_t(X=x)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}, \qquad X=0,\ 1,\ 2, 3, \ldots$$

مثال. بفرض أن عدد المكالمات الهاتفية التي تصل إلى مركز هاتفي معين تتبع توزيع بواسون بمعدل $\lambda=5$ في الثانية. أحسب احتمال وصول 7 مكالمات في ثانية ونصف.

$$t\lambda=1.5(5) \qquad P(X=7)=\frac{(1.5(5))^7 e^{-1.5(7)}}{7!}$$

(د) حساب احتمال عدد من الأحداث من فئة معينة.

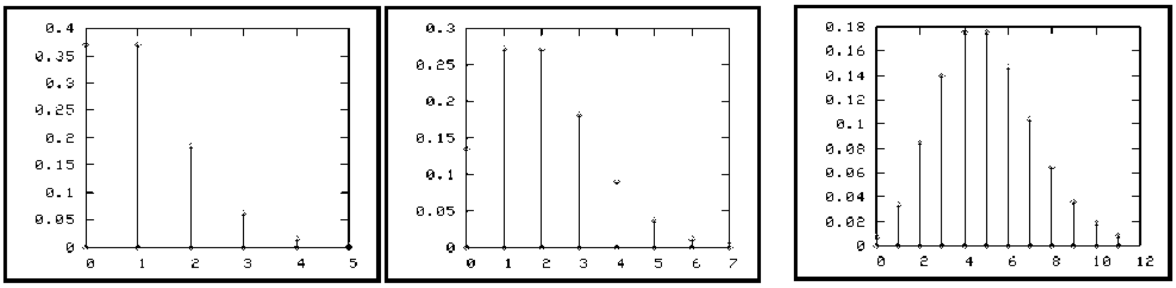
إذا كان X يتبع توزيع بواسون بمعدل λ، فإن $Y=aX$ هو الآخر يتبع توزيع بواسون بمعدل $a\lambda$.

مثال. بفرض أن عدد المكالمات الهاتفية التي تصل إلى مركز هاتفي معين تتبع توزيع بواسون بمعدل $\lambda=5$ في ثانية، وأن 6% من هذه المكالمات هي مكالمات دولية. أحسب احتمال أن تصل 9 مكالمات دولية في ثانية.

$$P(X=9)=\frac{(0.05(5))^9 e^{-0.05(5)}}{9!}$$

(هـ) التمثيل البياني لتوزيع بواسون

دالة توزيع بواسون هي دالة متناقصة لكون قوة e سالبة (وهي أقوى من قوة λ)، لكونه توزيعا متقطعا، يرسم توزيع بواسون من خلال مدرج أعمدة (Diagramme en bâtons). هذا قد يصعب ملاحظة سلوك التوزيع إلا باستخدام عدة أمثلة بمعالم متصاعدة بالتدريج، حيث يظهر أن التوزيع يقترب شيئا فشيئا من التوزيع الطبيعي لما λ كبير بما فيه الكفاية. الرسوم البيانية التالية تبين ذلك.



رسم 7 سلوك توزيع بواسون عند زيادة المعلمة من 1 إلى 2 إلى 5 (من اليسار إلى اليمين)

(و) استخدام توزيع بواسون بدلا من التوزيع الثنائي.

عندما $n\to\infty$ والمتوسط ثابت يؤول التوزيع الثنائي إلى التوزيع بواسون. عمليا يعطي توزيع بواسون نتائج قريبة من التوزيع الثنائي لما:

$n \geq 30$ و $np < 5$ أو $nq < 5$

ويستخدم بعض من الإحصائيين أيضا كشرط لاستعمال قانون بواسون بدلا من القانون الثنائي القاعدة التالية[1]:

$p \leq 0,1$ و $n \geq 25$

مثال : نأخذ عشوائيا 10 وحدات من انتاج آلة نسبة إنتاجها التالف 10 % . أحسب احتمال أن يكون هناك وحدتان تالفتان.

$$P(X = 2) = C^2_{10}\ (0,1^2)\ (0.9^8) = \mathbf{0.1937}$$

ط2. باستعمال توزيع بواسون: نحسب أولا قيمة المعلمة λ (معلمة قانون بواسون):

$$\lambda = \mu = np = 10 * 0,1 = 1$$

$$P(2) = \lambda^x * e^{-\lambda}/x! = (1^2 * e^{-1}) / 2! = 1/(2e) = \mathbf{1,1839}$$

(ز) الاستخدام العملي لتوزيع بواسون

ظل توزيع بواسون لفترة طويلة يستعمل فقط لتمثيل الأحداث النادرة، لكنه اليوم يستعمل في مجالات متعددة. فمن الدراسة الشهيرة ل (Ladislaus Bortkiewics) عن حوادث إصابات الجنود بصكات الجياد في الجيوش أصبح اليوم توزيع بواسون يستعمل في شتى المجالات؛ منها مراقبة الجودة إحصائيا، تسيير ظواهر الانتظار، الاتصالات (عدد المكالمات في وحدة زمن)، كما يستخدم في الفيزياء النووية لدراسة عدد الجزيئات المنبعثة من مادة مشعة وفي البيولوجيا الدقيقة (microbiologie) لمراقبة تكاثر البكتيريا في حقل تجارب، كما يستخدم في البيولوجيا وحتى في علم الأحوال الجوي.

في مجال التسيير، يستخدم توزيع بواسون بشكل خاص عند دراسة مسائل متعلقة "بظواهر الانتظار"؛ ففي هذا النوع من المسائل، كثيرا ما يفترض أن وصول الزبائن إلى مكان الخدمة يتبع توزيع بواسون. من أمثلة ذلك: عدد الطائرات التي تصل إلى المطار في وحدة زمن، عدد البواخر التي تصل إلى ميناء في وحدة زمن، عدد الزبائن الذين يصلون إلى مكتب بريدي في وحدة زمن، عدد المكالمات الهاتفية التي تصل إلى مركز هاتفي، عدد الحالات الاستعجالية التي تصل إلى مستشفى، ... تسمى هذه الظواهر في نظرية صفوف الانتظار "بظواهر الوصول".

مثال1. بينت دراسة أن عدد حوادث العمل في معمل معين يتبع توزيع بواسون بمعدل حادثتين يوميا.

أوجد احتمال أن لا يسجل أي حادث في يوم معين. أوجد احتمال حادث على الأقل في يوم:

$$P(X = 0) = \lambda^x * e^{-\lambda}/x! = \lambda^0 * e^{-\lambda}/0! \Rightarrow \mathbf{P(X = 0)\ =\ e^{-\lambda}} = e^{-2}$$

$$P(X \geq 1) = 1- P(0) = 1 - [\lambda^0 * e^{-\lambda}/0!] \Rightarrow \mathbf{P(X \geq 1) =\ 1 - e^{-\lambda}} = 1 - e^{-2}$$

مثال2. بينت دراسة إحصائية سابقة أن عدد السيارات التي تصل إلى محطة بنزين معينة بين الساعة 12:00 و12:05 هو في المتوسط 3 سيارات، كما بينت الدراسة أن عدد السيارات التي تصل إلى المحطة يتبع توزيع بواسون. أوجد احتمال أن تصل 4 سيارات بين 12:00 و12:05.

متوسط عدد السيارات في الساعة = 2 * 3 = 6 ومنه:

$$P(X=4) = 6^4 * e^{-6}/4! = 1296 * e^{-6}/24 = 54 * e^{-6}$$

في الأخير، ينبغي الإشارة إلى أن لتوزيع بواسون وأيضا للتوزيع الثنائي خصائص مهمة لا يتسع المقام لذكرها في إطار هذا الدرس، ولكن سنتعرض لبعضها في التطبيقات، لذلك نحيل الطالب إلى مطالعتها في المراجع المتخصصة؛ كما توجد توزيعات أخرى مهمة نظرا لتعدد استخداماتها مثل التوزيع المتماثل (distribution uniforme)، لم نتطرق لها في هذا الدرس، ندعو الطالب لاستكمالها من خلال بحثه الخاص.

[1] أنظر دروزبيك 1997، ص 262.

9 خلاصة

الجدول الملحق يلخص أهم النقاط حول التوزيعات المتقطعة الشهيرة.

التوزيع	متى يستخدم	القيم الممكنة للمتغيرة	الاحتمال	التوقع والتباين
الهندسي الزائد **X~ H(N, b,p)**	سحب بدون إرجاع. كريات من صنفين.	$X = \{0,1,2,\ldots,b\}$, $b \leq b + r = N$	$P(X=x)=\dfrac{C_b^x \cdot C_r^{n-x}}{C_N^n}$	$\mu = np,$ $\sigma^2 = npq\left(\dfrac{N-n}{N-1}\right)$ $q = r/N$ و $p = b/N$ n عدد الكريات المسحوبة N العدد الكلي للكريات b عدد الكريات البيضاء r ع الكريات الحمراء
الهندسي الزائد المتعدد	نفس شروط ت الهندسي الزائد مع وجود أكثر من صنفين من الكريات.	$X_i = \{0,1,2,\ldots,N_i\}$, $\Sigma x_i = n$, $\Sigma N_i = N$	$P(X_1=x_1,X_2=x_2,\ldots X_k=x_k)=$ $=\dfrac{C_{N1}^{x1}C_{N2}^{x2}C_{N3}^{x3}\cdots C_{Nk}^{xk}}{C_N^n}$	$E[X_i] = n\,(N_i/N) = np_i$
برنولي **X~B(1, p)**	تجربة واحدة (غير مكررة) تقبل نتيجتين.	$X = \{0, 1\}$	$P(X = 1) = p,$ $P(X = 0) = 1 - p = q$	$\mu = p, \quad \sigma^2 = pq$
الثنائي **X~B(n, p)**	تجارب ثنائية النتيجة، مكررة ومستقلة (p ثابت).	$X = \{0,1,2\ldots,n\}$	$P(X = x) = C^x_n\, p^x q^{n-x}$ $C_n^x = \dfrac{n!}{x!(n-x)!}$	$\mu = np$, $\sigma^2 = npq$
باسكال (الثنائي السالب)	X هي عدد التجارب اللازمة للحصول على عدد r من النجاحات في تجارب برنولية مكررة.	$X = \{r, r+1, r+2, \ldots, +\infty\}$	$P(X = x) = C^{r-1}_{x-1}\, p^r q^{x-r}$	$\mu = r/p$, $\sigma^2 = rq/p^2$
الهندسي	X هي عدد التجارب اللازمة للحصول على النجاح الأول في تجارب برنولية مكررة.	$X = \{1,2,\ldots,+\infty\}$	$P(X = x) = q^{x-1}p$	$\mu = 1/p,$ $\sigma^2 = q/p^2$
التوزيع المتعدد	هو تعميم للتوزيع الثنائي على تجربة مكررة متعددة النتائج.	$\forall i, 0 \leq x_i \leq N_i,$ $\sum_{i=1}^{k} x_i = n,$ $\sum_{i=1}^{k} N_i = N$	$P(X_1=x_1, X_2=x_2, \ldots, X_k=x_k) = \dfrac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$	$E(X_k) = np_k$ $V(X_k) = np_k q_k$
بواسون **X~P(λ)** **$\lambda > 0$**	X عدد النجاحات في عدد كبير جدا من التجارب البرنولية (عدد الوحدات التالفة في شحنة). أو أيضا عدد من الأحداث في فترة زمن.	$X = \{0,1,2,\ldots,+\infty\}$	$P(X=x)=\dfrac{\lambda^x e^{-\lambda}}{x!}$ $P(X = 0) = e^{-\lambda}$ $P(X \geq 1) = 1 - e^{-\lambda}$	$E(x) = V(x) = \lambda$

المبحث 2. التوزيعات الاحتمالية الشائعة المستمرة

التوزيع الطبيعي
التوزيع الأسي
توزيع قاما
توزيع بيتا

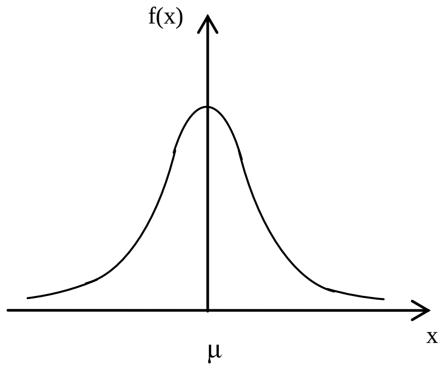
1 التوزيع الطبيعي أو توزيع لابلاس قوس[1] D. Normale ou D. de Laplace -Gausse

يعد التوزيع الطبيعي من أهم التوزيعات الاحتمالية شائعة الاستخدام لما له من خصائص تنطبق على نسبة كبيرة من الظواهر الطبيعية والاجتماعية والاقتصادية. فلو اخترنا بالصدفة مئة أو ألفا من المارين في شار ع ما وقسنا أطوالهم لوجدنا نسبة كبيرة منها قريبة من متوسط ما، ونسبة قليلة من طوال القامة ونسبة مقاربة لها من قصار القامة. ومثل هذا بالنسبة للأوزان. ولو مثلنا هذه البيانات في معلم متعامد متجانس لكان المنحنى الذي يمثل النسبة، أو ما يمكن أن نسميه الاحتمال، ذا شكل جرسي متماثل حول المتوسط وهي صفات التوزيع الطبيعي (الشكل9) :

(أ) صيغة القانون

تكتب دالة الكثافة لمنحنى للتوزيع الطبيعي كما يلي:

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty}$$



رسم**8 الشكل العام للتوزيع الطبيعي**

حيث μ و σ هما على التوالي التوقع والانحراف المعياري. ونكتب $X \sim N(\mu, \sigma)$

دالة التوزيع (الدالة التجميعية) للتوزيع الطبيعي تكتب كما يلي:

$$F(x) = P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-1/2\left(\frac{v-\mu}{\sigma}\right)^2} dv$$

المتغيرة المركزية أو المعيارية : تستخدم المتغيرة المعيارية $Z = (X-\mu)/\sigma$ لتكوين الجداول الإحصائية للاحتمالات: $P(0 \le Z \le z)$ أو $F(z) = P(Z \le z)$،

حيث تسمح بكتابة الدالة f وF بدلالة مجهول واحد Z بدلا من 3 مجاهيل x وμ و σ وذلك كما يلي:

$$\boxed{f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty}$$

---

[1] باسم العالمان الرياضيان الفيزيائيان و الفلكيان الفرنسي Pière Simon de Laplace (1749-1827) والألماني Carl Freidrich Gauss - الصورة لهذا الأخير-، (1777-1855) الذين كانا من أوائل من اكتشف هذا القانون. أما من أعطاه تسمية التوزيع الطبيعي فهو Pearson في 1893. أنظر J. J. Droesbeke (1997)، ص329 .

$$F(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-u^2}{2}} du$$

بالنظر إلى العلاقة الخطية بين المتغيرتين X و Z، فإن Z تتبع نفس توزيع X أي التوزيع الطبيعي. ونعلم أن:

$$\mathbf{E(Z) = 0} \qquad \mathbf{V(Z) = 1}$$

(ب) خصائص التوزيع الطبيعي

الدالة المتجددة للعزوم :

$$M_x(t) = \exp\left(t\mu + \frac{t^2\mu^2}{2}\right)$$

من خصائص التوزيع الطبيعي أنه يعتبر معتدلا لا مدببا ولا مفلطحا، حيث يعتبر معامل التفلطح $\alpha_4 = 3$ للتوزيع الطبيعي معيارا لاعتدال المنحنيات.

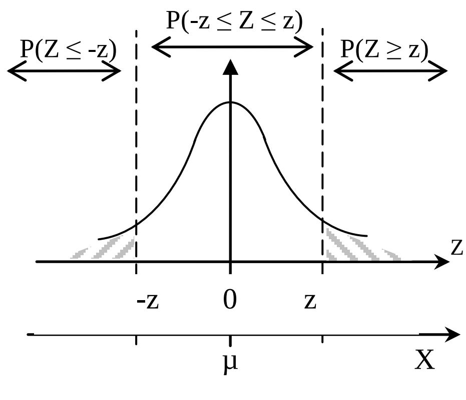
من خصائص التوزيع الطبيعي أيضا أنه متماثل حول $\alpha_3 = \frac{\mu_3}{\sigma^3} = 0$ القيمة المتوقعة

تماثل منحنى X حول المتوسط (أنظر الشكل 3) يعني تماثل لمنحنى Z حول 0، مما يعني أنه من أجل أي قيمة للمتغيرة المعيارية

: $z > 0$

$$P(0 \le Z \le z) = P(-z \le Z \le 0) = P(-z \le Z \le z) / 2$$

$$P(Z \le -z) = 1 - P(Z \le z) = P(Z \ge z)$$



رسم 9 استخدام تماثل الوزيع الطبيعي في حساب الاحتمالات

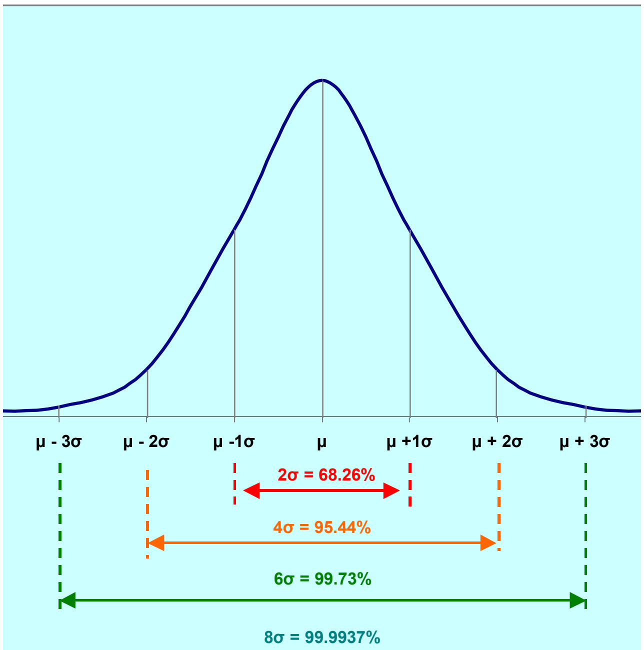
و لقد تم باستخدام المتغيرة المعيارية Z حساب الاحتمالات (المساحات) تحت المنحنى ومنها خاصة:

$P(-\sigma \le X \le \sigma) = P(-\sigma \le Z \le \sigma) = 0.6837,$

$P(-2\sigma \le X \le 2\sigma) = P(-2 \le Z \le 2) = 0.9544,$

$P(-3\sigma \le X \le 3\sigma) = P(-3 \le Z \le 3) = 0.9973.$

هذه القيم وغيرها متوفرة في الجداول الإحصائية التي نجدها في الكثير من المراجع، كما يمكن حسابها باستخدام الحاسوب.



رسم 10 المساحات الأساسية تحت منحنى التوزيع الطبيعي

مثال: باستعمال الجداول الاحصائية 1) أحسب : $P(0 \le Z \le z)$ حيث $z = 1, 2, 3$

2) أحسب $P(-z \le Z \le z)$ من أجل نفس القيم ل z.

1) 0.3413 ، 0.47725 ، 0.49865

2) 0.6827، 0.9545، 0.9973

(ج) العلاقة بين التوزيع الطبيعي والتوزيع الثنائي

في حالة n كبيرة وp غير قريب من 0 يمكن اعتبار التوزيع الثنائي كتقريب جيد للتوزيع الطبيعي. ويعطي التوزيعان نتائج أكثر تقاربا كلما كانت n كبيرة أكثر. ونكتب:

$$\lim_{n \longrightarrow \infty} P(a \le z \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz, \qquad z = \frac{x - np}{\sqrt{npq}}$$

ويسرع تقارب التوزيع الثنائي من التوزيع الطبيعي كون p قريب من 0.5.

قاعدة التقريب:

- **عموما نعتبر التقريب إلى التوزيع الثنائي ملائما عندما np و nq كلاهما أكبر من 5.**
- عدد من الاحصائيين[1] يعتمد قاعدة أخرى هي أن يكون أحد الشرطين التاليين متوفرين:

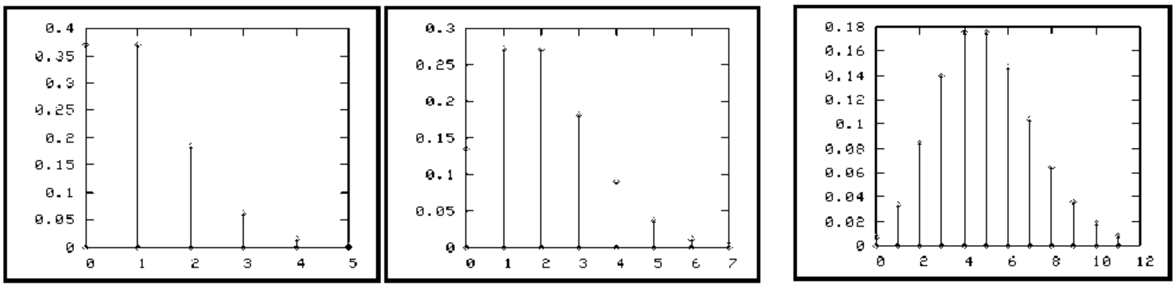
$$npq \ge 9 \qquad ou \qquad n \ge 20, np \ge 10, nq \ge 10$$

[1] أنظر دراوزبيك 1997. ص 262.

(د) العلاقة بين التوزيع الطبيعي وتوزيع بواسون

عندما $\lambda \to \infty$ فإن التوزيعين الطبيعي وبواسون يعطيان نتائج متطابقة . ونكتب:

$$\lim_{\lambda \longrightarrow \infty} P\left(a \le \frac{x-\lambda}{\sqrt{\lambda}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-z^2}{2}} dz$$



رسم 11 سلوك توزيع بواسون عند زيادة المعلمة من 1 إلى 2 إلى 5 (من اليسار إلى اليمين)

قاعدة التقريب:

- **عموما نعتبر أن التقريب ملائم من التوزيع بواسون إلى التوزيع الطبيعي عندما $10 \le \lambda$**
- **فيما يعتمد عدد من الإحصائيين[1] كشرط للتقريب $\lambda \ge 15$**

ويمكن أن تتقارب نتائج التوزيعات الثلاث معا: الثنائي، بواسون والطبيعي حسب الشروط المذكورة (أنظر حل التطبيق أدناه.)

2 التوزيع الأسي Distribution exponentielle

عادة ما يستخدم التوزيع الأسي في مسائل متعلقة بقياس الزمن. من ذلك مدة خدمة شباك البريد، مدة مكالمة هاتفية، مدة تفريغ باخرة شحن، مدة تصليح آلة، مدة انتظار زبون قبل الحصول على الخدمة...في العلوم الدقيقة يستخدم التوزيع الأسي لتمثيل مدة حياة الذرات المشعة (atomes radioactives) قبل أن تتفكك، حيث يعبر الوسيط عن اللحظة التي يبقى فيها نصف المجتمع الأصلي[2].

من الضروري فهم الآتي: كقاعدة عامة يستخدم التوزيع الأسي لتمثيل مدة حياة ظاهرة ما إذا كان لها متوسط ثابت $1/\lambda$ وكانت هذه الظاهرة لا تخضع للتقادم (vieillissement) أي أن مدة حياة الظاهرة بعد لحظة ما T لا تتبع اللحظة T؛ أي لا تتأثر بالمدة التي دامتها الظاهرة من قبل. مثلا قد نستبعد استخدام التوزيع الأسي لتمثيل مدة حياة آلة عاملة قبل تعطلها لأن احتمال تعطلها في لحظة ليس مستقلا عن المدة التي عملتها الآلة من قبل، كذلك الأمر بالنسبة لمدة حياة الإنسان.

عمليا، نتحقق من دقة تمثيل التوزيع الأسي _أو أي توزيع آخر_ لظاهرة ما من خلال تقنيات اختبارات الفروض، وبالتحديد اختبار التجانس و التعديل.

نشير أخيرا إلى أن للتوزيع الأسي علاقة بالتوزيع بواسون، فإذا كان وقوع أحداث ما يتبع هذا التوزيع، فإن المدة بين وقوع حدثين تتبع التوزيع الأسي؛ كمثال على ذلك، إذا كان وصول الزبائن إلى مركز خدمة ما يتبع التوزيع بواسون فإن المدة الزمنية بين وصول زبون "أ" والزبون الموالي تتبع التوزيع الأسي. تتبين هذه العلاقة عند استنتاج صيغة القانون الأسي.

---

[1] المرجع نفسه.

[2] راجع موقع موسوعة Wikipédia .

(أ) صيغة القانون الأسي أو دالة الكثافة و الدالة التجميعية للتوزيع.

بينت دراسة أن عدد حوادث العمل في معمل معين تتبع توزيع بواسون بمعدل λ حادث يوميا.

أوجد احتمال أن يسجل حادث على الأقل (حادث أو أكثر) في مدة t يوم.

$P(X \geq 1) = 1 - P(0) = 1 - [\lambda^{0t} * e^{-\lambda t}/0!] \quad \Rightarrow \mathbf{P(X \geq 1) = 1 - e^{-\lambda t}}$

لنرمز ب T للزمن (باليوم) بين حادثين إذن سيكون لدينا f(t) دالة الكثافة للزمن بين حادثين، و $F(t) = P(T \leq t)$ دالة التوزيع ل T.

لنحسب احتمال P أن يكون الزمن بين حادثين يوم أو أقل:

لدينا $P = P(T \leq t = 1)$ إذن:

$P = F(t = 1)$ .... (1)

لاحظ من ناحية أخرى أن P هو معادل لاحتمال أن يسجل على الأقل حادث في يوم معين:

$P = P(X \geq 1) = 1 - e^{-\lambda t}$..........(2)

من (1) و(2) نستنتج أن $\boxed{\mathbf{F(t) = 1 - e^{-\lambda t}}}$...........(3)

و منه $f(t) = F(t)' = (1 - e^{-\lambda t})'$

إذن $\mathbf{f(t) = \lambda e^{-\lambda t}}$

قاعدة: إذا كان حدث عشوائي ما يتكرر في الزمن وفق توزيع بواسون:

$$\boxed{p_\tau(x) = \frac{(\lambda\tau)^x e^{-\lambda\tau}}{x!}}$$

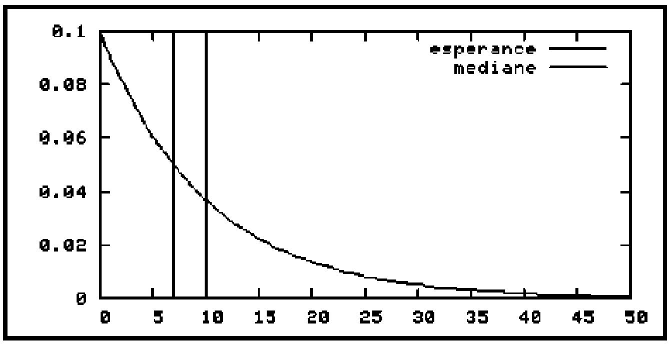
فإن الزمن T بين حادثين يتبع التوزيع التالي:

$$f(\tau) = \begin{cases} \lambda e^{-\lambda\tau} & , \quad \tau > 0 \\ 0 & , \quad \tau \leq 0 \end{cases}$$

حيث λ عدد حقيقي موجب.

و يسمى هذا التوزيع التوزيع الأسي ويسمى أيضا التوزيع الأسي السالب لعلاقته بتوزيع بواسون.

(ب) التمثيل البياني للتوزيع الأسي



رسم 12 دالة الكثافة للتوزيع الأسي

(ج) خصائص التوزيع الأسي

$$\mu = 1/\lambda \quad , \quad \sigma^2 = 1/\lambda^2 \quad , \quad Med = \mu \ln(2) < \mu \quad , \quad M_x(t) = \frac{\lambda}{\lambda - t}$$

3 توزيع قاما Distribution gamma

توزيعي قاما و بيتا يمثلان مجموعة واسعة من التوزيعات ذات معلمتين تتميز بمرونة وقدرة على توليد توزيعات متعددة حسب قيم المعلمتين. ندرس هذين التوزيعين أيضا لعلاقتهما بالتوزيعات F، t، و كا2. يستخدم توزيع قاما لتمثيل بعض الظواهر مثل توزيع الدخل والادخار تحت شروط معينة[1].

(أ) **صيغة القانون**.

نقول عن متغيرة عشوائية أنها تتبع توزيع قاما إذا كانت دالة كثافتها كما يلي:

$$f(x)=\begin{cases}\dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x>0\\ 0 \quad , & x\le 0\end{cases} \quad , \alpha>0, \beta>0$$

حيث $\Gamma(\alpha)$ هي الدالة قاما: $\Gamma(\alpha)=\int_0^{\infty}t^{\alpha-1}e^{-t}dt \qquad \alpha>0$

ونكتب $X \sim \Gamma(\alpha, \beta)$

(ب) **خصائص توزيع قاما**

$$\boldsymbol{\mu = \alpha\beta \quad , \quad \sigma^2 = \alpha\beta^2, \quad M(t) = (1-\beta t)^{-\alpha}}$$

Pour $\alpha>1$: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ et si $\alpha \in N$: $\Gamma(\alpha) = (\alpha-1)!$, $\Gamma(1/2) = \sqrt{\pi}$

من خصائص توزيع قاما علاقته بالتوزيع الأسي كما سنرى في السلسلة.

مثال. أحسب ما يلي:

$$\int_0^{\infty}t^4e^{-t}dt, \qquad \int_0^{\infty}x^6e^{-x}dx, \qquad \int_0^{\infty}\frac{e^{-x}}{x^{1/2}}dx, \qquad \Gamma(7), \quad \Gamma(4.5), \qquad \Gamma(2.5).$$

$$\int_0^{\infty}t^4e^{-t}dt=\Gamma(5)=4!=24, \int_0^{\infty}x^6e^{-x}dx=\Gamma(7)=6!=720, \int_0^{\infty}\frac{e^{-x}}{x^{1/2}}dx=\int_0^{\infty}x^{-1/2}e^{-x}dx=\Gamma(1/2)=\sqrt{\pi}$$

$$\Gamma(4.5)=\Gamma(3.5+1)=3.5\Gamma(3.5)=3.5(2.5)\Gamma(2.5)=3.5(2.5)(1.5)(0.5)\Gamma(0.5)=3.5(2.5)(1.5)(0.5)\sqrt{\pi}$$

$$\Gamma(7)=6!=720, \qquad \Gamma(2.5)=1.5\Gamma(0.5)=1.5\sqrt{\pi}$$

مثال2. أحسب المتوسط والتباين للمتغيرات العشوائية X و Y و Z المعرفة كما يلي:

$$f(x)=\begin{cases}\dfrac{x^4e^{-x/2}}{2^5\Gamma(5)}, & x>0\\ 0 \quad , & x\le 0\end{cases}, \quad f(y)=\begin{cases}\dfrac{y^3e^{-y/4}}{4^4(6)}, & y>0\\ 0 \quad , & y\le 0\end{cases}, \quad f(z)=\begin{cases}\dfrac{z^2e^{-z}}{6}, & z>0\\ 0 \quad , & z\le 0\end{cases}$$

$$\mu_x=\alpha\beta=5(2)=10, \sigma^2{}_x=\alpha\beta^2=5(2^2)=20, \mu_y=4(4)=16, \sigma^2{}_y=4(4^2)=64, \mu_z=3(1)=3, \sigma^2{}_z=3$$

[1] أنظر: آيفازيان و آخرون، مبادئ النمذجة و المعالجة الأولية للبيانات، سلسلة : Mathématiques، Editions Mir، موسكو، 1983، ترجمه من الروسية إلى الفرنسية جيلالي مبارك، 1986. ص158.

4 توزيع بيتا Distribution bêta

يتميز توزيع بيتا بمرونته الكبيرة تبعا لقيم معلمتيه (أنظر الرسم 14) حيث يستخدم لحساب توزيع t^2، F، التوزيع الثنائي، الثنائي السالب وغيرها[1]، وتستخدم لتمثيل بعض المتغيرات التي تتراوح بين 0 و 1، مثل نسبة ما كنسبة التالف أو المبيعات، إلخ.

(أ) **صيغة القانون**.

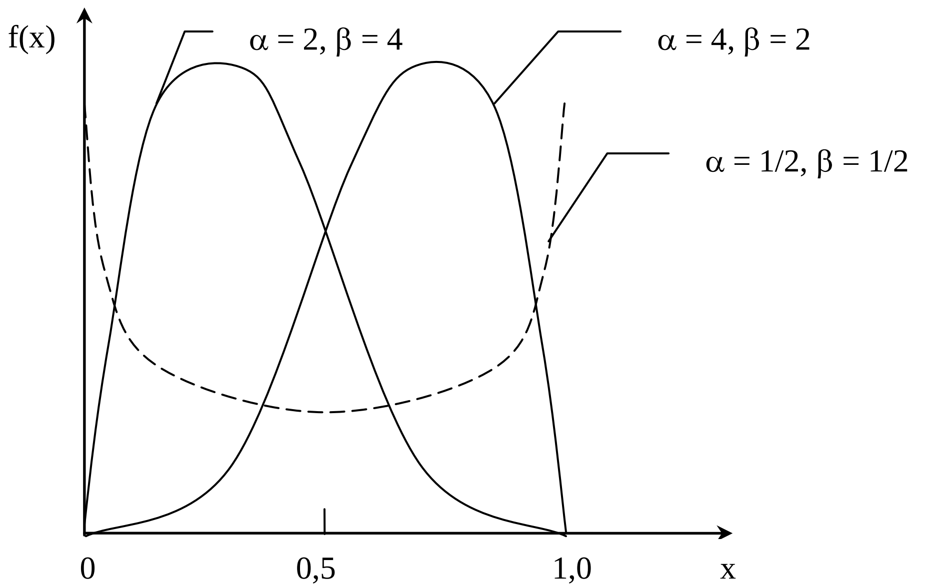
نقول عن متغيرة عشوائية أنها تتبع توزيع بيتا إذا كانت دالة كثافتها كما يلي:

$$\boxed{f(x)=\begin{cases}\dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)} & 0<x<1 \qquad (\alpha,\beta>0)\\ 0 & ailleurs\end{cases}}$$

حيث $B(\alpha, \beta)$ هي الدالة بيتا: $B(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du, \qquad \alpha,\beta>0$

و نكتب $X \sim B(\alpha, \beta)$

$\alpha = 4, \beta = 2$



رسم 13 التمثيل البياني لدالة الكثافة للتوزيع بيتا من أجل قيم مختلفة للمعالم

(ب) **خصائص توزيع بيتا**

$$\alpha=\frac{\alpha}{\alpha+\beta}, \qquad \sigma^2=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

[1] المرجع السابق.

ج) **العلاقة بين الدالتين قاما وبيتا:**

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

مثال. أحسب ما يلي:

$$B(3,4),\quad B(1/2,1/2),\quad B(n,2),\quad B(1,n),\quad B(n,1)\quad (n\in N)$$

$$B(3,4)=\frac{(3-1)!(4-1)!}{(7-1)!}=\frac{2}{120}=\frac{1}{60},\quad B(1/2,1/2)=\frac{(\sqrt{\pi})(\sqrt{\pi})}{(1/2)+(1/2)}=\pi,$$

$$B(n,\ 2)=\frac{(n-1)!1!}{(2+1)!}=\frac{1}{n(n+1)!},\qquad B(1,n)=\frac{1(n-1)!}{n!}=\frac{(n-1)!}{n(n-1)!}=\frac{1}{n},\qquad B(n,1)=\frac{(n-1)!1}{n!}=\frac{1}{n}$$

مثال2. أحسب ما يلي: $\int_0^1 x(1-x)dx$، $\int_0^1 x^4(1-x)^3dx$، $B(3,2)$

$$B(3,2)=1/[n(n+1)]=1/[3(4)]=1/12$$

$$\int_0^1 x^4(1-x)^3dx=B(5,4)=\frac{\Gamma(5)\Gamma(4)}{\Gamma(5+4)}=\frac{4!3!}{8!}=\frac{1}{280},\qquad \int_0^1 x(1-x)dx=B(2,2)=\frac{1!1!}{3!}=1/6$$

وباستعمال العلاقة $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ نجد أن دالة الكثافة للتوزيع بيتا تكتب أيضا:

$$f(x)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, & 0<x<1\\ 0\qquad , & ailleurs\end{cases}$$

للتوزيعين قاما وبيتا علاقة بعدد من التوزيعات المهمة كالتوزيع الأسي وتوزيع كاي تربيع، من ذلك **مثلا** أن التوزيع الأسي هو حالة خاصة من توزيع قاما عندما **$\alpha = 1$, $\beta = 1/\lambda$**.

مثال3. أحسب النسبة المتوقعة للإنتاج التالف والتباين، إذا كانت نسبة الإنتاج التالف تتبع التوزيع التالي:

$$f(x)=\begin{cases}6(1-x)^5, & 0<x<1\\ 0\quad , & \text{sinon}\end{cases}$$

من المثال السابق لدينا: $B(1,n)=1/n\Rightarrow 6=1/B(1,6).$

بوضع α و β يساويان 1 و 6 على التوالي، نجد أن $X\sim B(1,6)$

ومنه: $\mu=\dfrac{\alpha}{\alpha+\beta}=1/2,\qquad \sigma^2=\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=\dfrac{4}{16(5)}=1/20.$

مثال4. نسبة الإنتاج المباع في مؤسسة تتبع التوزيع التالي. أحسب النسبة المتوقعة، واحتمال أن تبلغ النسبة أكثر من 35%.

$$f(x)=\begin{cases}12x^2(1-x), & 0<x<1\\ 0 & ,\ \text{sinon}\end{cases}$$

$$12=3*4\Rightarrow 12=1/B(3,2)\Rightarrow X\sim B(3,2)\Rightarrow \mu=\frac{\alpha}{\alpha+\beta}=\frac{3}{3+2}=3/5=60\%$$

$$P(X>0.35)=\int_{0,35}^{1}12x^2(1-x)dx.$$

$$soit: v=x^2dx\ et\ u=1-x\Rightarrow P(X>0.35)=12\left(\left[(1-x)\frac{x^3}{3}\right]_{0,35}^{1}+\int_{0,35}^{1}x^2dx\right)=0.3125$$

5 خلاصة

الجدول التالي يلخص خصائص التوزيعات الاحتمالية المستمرة الأكثر استخداما.

التوزيع	دالة الكثافة ، التوقع والتباين		خصائص التوزيع
التوزيع الطبيعي المعياري X~N(0, 1)	$\boxed{f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\qquad -\infty<z<\infty}$ E(Z) = 0, V(Z) = 1		$P(Z\le -z)=1-P(Z\le z)=P(Z\ge z)$ $P(-\sigma\le X\le\sigma)=P(-1\le Z\le 1)=0.6837$, $P(-2\sigma\le X\le 2\sigma)=P(-2\le Z\le 2)=0.9544$, $P(-3\sigma\le X\le 3\sigma)=P(-3\le Z\le 3)=0.9973$.
التوزيع الأسي	$f(\tau)=\begin{cases}\lambda e^{-\lambda\tau}, & \tau>0\\ 0, & \tau\le 0\end{cases}$	$\mu=1/\lambda$, $\sigma^2=1/\lambda^2$ $F(x)=1-e^{-\lambda x}$	$P(X\le\mu)=0.63$
توزيع قاما X~Γ(α,β)	$f(x)=\begin{cases}\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}, & x>0\\ 0, & x\le 0\end{cases}\quad ,\ \alpha>0,\beta>0$ $\mu=\alpha\beta$, $\sigma^2=\alpha\beta^2$		$\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt\qquad \alpha>0$
توزيع بيتا X~ B(α, β)	$f(x)=\begin{cases}\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)} & 0<x<1\\ 0 & ailleurs\end{cases}\quad(\alpha,\beta>0)$ $\alpha=\frac{\alpha}{\alpha+\beta},\quad \sigma^2=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$		$B(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du,\qquad \alpha,\beta>0$ $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

الفصل V. المتغيرات العشوائية متعددة الأبعاد

المتغيرة الثنائية الاستقلال التباين والارتباط

درسنا في الفصل الأول المفاهيم والقواعد الأساسية في نظرية الاحتمالات كالاحتمال البسيط والاحتمال المتعدد وقاعدة جمع وضرب الاحتمالات. تساعد هذه المفاهيم المسير على التقدير وعلى اتخاذ القرار المناسب تجاه مسائل متشعبة وغير مؤكدة النتائج. في الفصل الثاني درسنا مفهوم المتغيرة العشوائية والتوزيع الاحتمالي (أو القانون الاحتمالي) وتطرقنا إلى عدد من التوزيعات الاحتمالية الشهيرة. تستخدم المتغيرة العشوائية لتمثيل الظواهر المختلفة من أجل دراستها والتوقع بشأنها. في هذا الفصل سنتناول نوعا من المتغيرات العشوائية وتوزيعاتها وهي المتغيرات العشوائية ذات أكثر من بعد، بالإضافة إلى مفاهيم أخرى مهمة متعلقة بالارتباط بين المتغيرات. ذلك أن العديد من الظواهر والمسائل التي تطرح أمام المسير تتضمن أكثر من متغيرة واحدة، فنتيجة نشاط مؤسسة هي محصلة العائد والتكاليف، ومحصول موسم زراعي يتأثر بمتغيرات عدة مثل كمية الأمطار والأسمدة والمساحة المزروعة. سوف نقتصر في دراستنا على المتغيرة ذات بعدين اثنين وهو ما يطلق عليه المتغيرة الثنائية.

المبحث 1. المتغيرة الثنائية

التوزيعات المشتركة المتقطعة والدوال الحدية (الهامشية)
التوزيعات المشتركة المتصلة
التوزيع الشرطي

1 التوزيعات المشتركة المتقطعة والدالة الهامشية (الحدية) Fonction marginale

أ) تعريف

المتغيرة الثنائية هي متغيرة تتوقف ليس على قيمة واحدة هي قيمة X مثلا و إنما تتوقف على قيمة متغيرتين اثنتين. مثال ذلك، معدل الطالب يتوقف على نقطة الرقابة المستمرة و نقطة التطبيق أو نقطة السداسي الأول ونقطة السداسي الثاني. كذلك نتيجة السنة المالية تتوقف على متغيرتي التكاليف و الإيرادات، وهكذا. التعريف الدقيق للمتغيرة الثنائية يتأتى باستخدام الترميز كما يلي:

لتكن لدينا متغيرتان عشوائيتان متقطعتان X وY، لنرمز للاحتمال: $P(X = x, Y = y)$ ب $f(x,y)$:

$$f(x,y) = P(X = x, Y = y)$$

$$f(x,y) \geq 0$$

$$\sum_x \sum_y f(x,y) = 1$$

تسمى الثنائية (X,Y) متغيرة ذات بعدين و$f(x,y)$ دالة الكثافة الاحتمالية لها ونقول أيضا دالة الكثافة المشتركة للمتغيرتين X وY ويمكن التعبير عنها عن طريق جدول للاحتمالات المشتركة (جدول التوزيع المشترك).

Y \ X	y_1	y_2	. . .	y_m	$f_1(x)$
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	. . .	$f(x_1,y_m)$	$f_1(x_1)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	. . .	$f(x_2,y_m)$	$f_1(x_2)$
. . .	. . .	. . .	. . .	. . .	. . .
x_n	$f(x_n,y_1)$	$f(x_n,y_2)$	. . .	$f(x_n,y_m)$	$f_1(x_n)$
$f_2(y)$	$f_2(y_1)$	$f_2(y_2)$	. . .	$f_2(y_n)$	1

احتمال $X = x$ يحسب ويكتب كما يلي $P\ (X = x) = f_1(x) = \sum_{k=1}^{m} f(x,y_k)$

احتمال $Y = y$ يحسب و يكتب كما يلي $P\ (Y = y) = f_2(y) = \sum_{i=1}^{n} f(x_i,y)$

الدالتين $f_1(x)$ و$f_2(y)$ تسميان الدالتان الهامشيتان (الحديتان) حيث : $\sum f_1(x) = 1$ و$\sum f_2(y) = 1$

ب) الدالة التجميعية

الدالة التجميعية للمتغيرة الثنائية (X,Y) تكتب كما يلي:

$$F\ (x, y) = P\ (X \le x, Y \le y) = \sum_{u\le x} \sum_{v\le y} f\ (u, v)$$

مثال: نرمي قطعة نقدية وحجر نرد، نرمز بX لعدد مرات ظهور الصورة، وY للرقم الذي يظهر من مكعب النرد.

- أكتب التوزيع الاحتمالي المشترك للمتغيرتين،
- أحسب احتمالات الأحداث التالية: الحصول على صورة مع الرقم 6، الحصول على الصورة، الحصول على الرقم 6.
- أحسب الاحتمال $P(X \le 1, Y \le 3)$ ، $P(X \le 2, Y \le 6)$.

X\Y	1	2	3	4	5	6	$f_1(x)$
0	1/12	1/12	1/12	1/12	1/12	1/12	/2
1	1/12	1/12	1/12	1/12	1/12	**1/12**	**1/2**
$f_2(y)$	1/6	1/6	1/6	1/6	1/6	1/6	1

احتمال الحصول على صورة مع الرقم 6 : $P\ (X = 1 \text{ et } Y = 6) = f(1, 6) = 1/12$

احتمال الحصول على الصورة $P\ (X = 1) = f_1(1) = \sum_{k=1}^{n} f(1, y_k) = 1/12 + 1/12 + . . . = 1/2$

احتمال الحصول على الرقم 6 $P\ (Y = 6) = f_2(6) = \sum_{i=1}^{m} f(x_i,6) = 1/12 + 1/12 = 1/6$

$$P(X \le 1, Y \le 3) = F(1, 3) = \sum_{u\le 1} \sum_{v\le 3} f(u,v) = 6(1/12) = 1/2\ ,\ P(X \le 2, Y \le 6) = 1$$

سؤال. من بين التوزيعات الاحتمالية الشهيرة التي رأينا في الفصل الثاني أيها يعتبر توزيعا مشتركا؟ (الجواب: التوزيع المتعدد.)

2 التوزيعات المشتركة المتصلة

(أ) تعريف

لتكن لدينا X وY متغيرتان ع متصلتان، نعرف دالة الكثافة الاحتمالية المشتركة لهما[1] كما يلي:

$$P(X = x, Y = y) = f(x, y)$$

$$f(x, y) \geq 0$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1$$

(ب) الدالة التجميعية

نكتب دالة التوزيع (الدالة التجميعية) كما يلي:

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{y} f(u, v)\, du\, dv$$

و يمكن استنتاج دالة الكثافة المشتركة من الدالة التراكمية بالاشتقاق كما يلي: $f(x, y) = \partial^2 F / (\partial x\, \partial y)$

من جهة أخرى $F_1(x) = P(X \leq x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} f(u, v)\, du\, dv$

$$F_2(y) = P(Y \leq y) = \int_{u=-\infty}^{\infty}\int_{v=-\infty}^{y} f(u, v)\, du\, dv$$

ونسمي الدالتان $F_1(x)$ و$F_2(y)$ الدالتان التراكميتان (التجميعيتان) الهامشيتان (الحديتان).

ولتحديد احتمال X وY محصورتان في مجالين ما نكتب:

$$P(a \leq x \leq b, c \leq y \leq d) = \int_{x=a}^{b}\int_{y=c}^{d} f(x, y)\, dx\, dy$$

(ج) الدوال الهامشية

الدالتان الهامشيتان (الحديتان) للكثافة الاحتمالية للثنائية (X, Y) فيعبر عنها كما يلي:

$$f_1(x) = \int_{v=-\infty}^{+\infty} f(x, v)\, dv\,, \qquad f_2(y) = \int_{u=-\infty}^{+\infty} f(u, y)\, du$$

مثال: لدينا دالة الكثافة الاحتمالية المشتركة للمتغيرتين X وY المعرفة كما يلي:

$$f(x, y) = \begin{cases} \dfrac{xy}{96} & , \quad 0 < x < 4,\ 1 < y < 5 \\ 0 & \text{sinon} \end{cases}$$

أكتب دالة التوزيع الهامشية لكل من المتغيرتين أحسب احتمال 2 > x > 0 ، أحسب احتمال 3 > y > 1.

$$F_1(x) = P(X \leq x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} f(u, v)\, du\, dv$$

$$* \ x < 0 : F_1(x) = 0,$$

[1] ونقول أيضا دالة الكثافة الاحتمالية للثنائية (X,Y) .

* $0 \le x < 4$:

$F_1(x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} uv/96\ du\ dv = 0 + \int_{u=0}^{x}\int_{v=1}^{5} uv/96\ du\ dv$

$= 1/96 \int_{u=0}^{x}[\int_{v=1}^{5} uv\ dv]\ du = 1/96\int_{u=0}^{x}[12u]\ du = x^2/2\ (12/96) = \mathbf{x^2/16}$.

* $x \ge 4$: $F_1(x) = 1$

$$F_1(x) = \begin{cases} 0 & x < 0 \\ x^2/16 & 0 \le x < 4 \\ 1 & x \ge 4 \end{cases}$$

$F_2(y) = P(Y \le y) = \int_{u=-\infty}^{\infty}\int_{v=-\infty}^{y} f(u, v)\ du\ dv$

Pour $y < 1$: $F_2(y) = 0$,

* $1 \le y < 5$:

$F_2(y) = 0 + \int_{u=0}^{4}\int_{v=1}^{y} uv/96\ dudv = 1/96 \int_{u=0}^{4}[\int_{v=1}^{y} uv\ dv]\ du = 1/96 \int_{u=0}^{4}[u\,(y^2 - 1) / 2]\ du$

$= (1/2 * 1/96)\,(y^2 - 1)\,(u^2/2)_0^4 = (1/(2*96))\,(y^2 - 1)\,(16/2) = (y^2 - 1) / 24$

* $y \ge 5$: $F_2(y) = 1$

$$F_2(y) = \begin{cases} 0 & y < 1 \\ (y^2 - 1)/24 & 1 \le y < 5 \\ 1 & y \ge 5 \end{cases}$$

$P(0 < x < 2) = F_1(2) - F_1(0) = 4/16 = 1/4$,

$P(1 < y < 3) = 8 / 24 = 1/3$

3 التوزيع الشرطي Distribution conditionnelle

في حالة X، Y متغيرتان عشوائيتان متقطعتان، فإن دالة الكثافة الاحتمالية الشرطية ل $(X|Y = y)$ تكتب كما يلي $f(x/y)$ وتحسب كما يلي: $f(x/y) = \dfrac{f(x, y)}{f_1(x)} = \dfrac{P(X = x, Y = y)}{P(X = x)}$

و هذا استنادا إلى القانون التقليدي للاحتمالات الشرطية: $P(B/A) = \dfrac{P(A \cap B)}{P(A)}$

التوزيع الاحتمالي ل X حيث $Y = y$ هو مجموعة قيم المتغيرة X عند تثبيت Y والاحتمالات $f(x/y)$ المقابلة لها.

مثال. لتكن X عدد مرات الحصول على صورة عند رمي قطعة نقدية مرتين وY الفرق بالقيمة المطلقة بين عدد مرات الصورة وعدد مرات الكتابة.

أكتب التوزيع الاحتمالي لـ$Y|X = 1$، أكتب التوزيعان الاحتماليان ل $X|Y = 0$ و$X|Y = 2$.

X	0	1	2
P(X/Y=0)	0	1	0
P(X/Y=2)	1/2	0	1/2

Y	1	0
P(y/x=1)	0	1

في حالة X وYمتغيرتان ع متصلتان نكتب: $P(c \le Y \le d / x \le X \le x + dx) = \int_c^d f(y/x)\ dy$

4 خلاصة

احتمال ثنائية عشوائية و يحسب كما يلي: $f(x,y) = P(X = x, Y = y)$

للتعبير عن احتمال قيمة ما لإحدى المتغيرتين نكتب: $P(X = x) = f_1(x)$ و تسمى دالة الكثافة الهامشية.

الاحتمال التجميعي لمتغيرتين فيعبر عنه من خلال دالة التوزيع المشتركة: $F(x, y) = P(X \le x, Y \le y)$

تنطبق هذه التعاريف على كل من المتغيرات المتقطعة و المستمرة.

للتعبير عن التوزيع الاحتمالي لإحدى المتغيرتين بشرط أن تأخذ المتغيرة الثانية قيمة ما (0، مثلا) فنكتب:

$P(X/Y = 0)$

لحساب الاحتمالات الشرطية ل X تستخدم القاعدة: $P(X/Y = y) = \dfrac{f(x,y)}{f_1(x)} = \dfrac{P(X = x, Y = y)}{P(X = x)}$

المبحث 2. الاستقلال التباين والارتباط

1 تعريف استقلال متغيرتين

رأينا في الفصل الأول أن حدثين عشوائيين A وB يكونان مستقلان إذا كان:

$P(A \text{ et } B) = P(A)\,P(B)$

انطلاقا من هذه القاعدة، تكون المتغيرتان العشوائيتان المتقطعتان X وY مستقلتان إذا وفقط إذا كان:

$P(X = x, Y = y) = P(X = x)\,P(Y = y)$

$f(x, y) = f_1(x)\,f_2(y)$

في حالة كون المتغيرتين متصلتين نكتب: $P(X \le x, Y \le y) = P(X \le x)\,P(Y \le y)$

$F(x, y) = F_1(x)\,F_2(y)$

أي أن المتغيرتان المستقلتان هما اللتان يمكن كتابة دالة التوزيع المشتركة لهما (أو دالة الكثافة المشتركة) في شكل جداء دالتين هامشيتين تراكميتين (أو دالتين هامشيتين للكثافة).

مثال. ليكن X وY م ع مستمرين حيث دالة الكثافة المشتركة لهما معرفة كما يلي:

$$f(x,y) = \begin{cases} cxy, & 0 < x < 4,\ 1 < y < 5 \\ 0 & \text{sinon} \end{cases}$$

بين أن المتغيرتين X وY مستقلتين.

Soit $c = c_1 * c_2 \Rightarrow f(x, y) = c_1 c_2\, xy = c_1 x * c_2 y \Rightarrow f(x, y) = f_1(x) * f_2(y)$ cqfd

مثال2. ليكن X وY و Z م ع متقطعة. تمثل X عدد مرات الحصول على صورة في رمية لقطعة نقدية وY عدد مرات الحصول على صورة في رمية موالية. وZ الفرق بالقيمة المطلقة بين ’X و’Y اللذان يمثلان على التوالي عدد مرات الحصول على الصورة/الكتابة في مجموع رميتين لقطعة نقدية.

Y \ X	0	1
0	¼	¼
1	½	½

X’ \ Z	0	1	2
0	0	1	0
2	½	0	½

X’	0	1	2
p_x	¼	½	¼

Z	0	2
p_z	½	½

من الواضح أن X وY مستقلتان لأن $P(X = x,Y = y) = P(X = x)\ P(Y = y)$ عند كل قيم X وY.

على العكس من ذلك، نجد أن Z ليست مستقلة عن’X فمثلا:

$$P(X' = 0)\ P(Z = 2) = (1/4\ .\ 1/2) = 1/8 \quad \neq \quad P(X'= 0,\ Z = 2) = 1/2$$

2 توقع وتباين المتغيرة العشوائية متعددة الأبعاد

ينطبق كل من تعريف التوقع الرياضي والتباين الذين تناولناهما فيما سبق على المتغيرة العشوائية متعددة الأبعاد.

لتكن *X* و*Y* متغيرتان عشوائيتان متقطعتان، و*f(x, y)* دالة كثافة مشتركة لهما.

$$\mu_x = E(X) = \sum_x\sum_y x\, f(x, y) \quad , \quad \mu_y = E(Y) = \sum_x\sum_y y\, f(x, y)$$

$$\sigma^2_x = E[(x - \mu_x)^2] = \sum_x\sum_y (x - \mu_x)^2 f(x, y)\ ,\ \sigma^2_y = E[(y - \mu_y)^2] = \sum_x\sum_y (y - \mu_y)^2 f(x, y)$$

في حالة X و Y متغيرتان مستمرتان:

$$\mu_x = E(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, f(x, y)\, dx\, dy\ ,\ \mu_y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\, f(x,y)\, dx\, dy\ .$$

$$\sigma^2_x = E[(x - \mu_x)^2] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x, y)\, dx\, dy\ ,$$

$$\sigma^2_y = E[(y - \mu_y)^2] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y - \mu_y)^2 f(x, y)\, dx\, dy$$

مثال. ليكن لدينا التوزيع المشترك المعرف كما يلي:

المطلوب حساب:

$E(x), E(y), \sigma^2_y, \sigma^2_x$

X \ y	-4	-2	7
1	1/8	1/4	1/8
-5	1/4	1/8	1/8

$$E(x) = \sum_x\sum_y x\, f(x, y) = 1(1/8 + ¼ + 1/8) - 5(1/4 + 1/8 + 1/8) = 1/2 - 5/2 = -4/2 = -2$$

$$E(Y) = \sum_x\sum_y y\, f(x, y) = -4\ (1/8 + ¼) - 2\ (1/4 + 1/8) + 7\ (1/8 + 1/8) = -1/2$$

$$\sigma^2_x = E[(x - \mu_x)^2] = \sum_x\sum_y (x - \mu_x)^2 f(x, y)$$

$$= (1 + 2)^2\ (1/8 + ¼ + 1/8) + (-5 + 2)^2\ (1/4 + 1/8 + 1/8) = 9\ (1/2) + 9\ (1/2) = 9$$

$$\sigma^2_y = E[(y - \mu_y)^2] = \sum_x\sum_y (y - \mu_y)^2 f(x, y)$$

$$= (-4 + 1/2)^2\ (1/8 + 1/4) + (-2 + 1/2)^2\ (1/4 + 1/8) + (7 + 1/2)^2\ (1/8 + 1/8)$$

$$= 49/4\ (3/8) + 9/2\ (3/8) + (15/2)^2\ (2/8) = 651 / 32 = 20,34$$

كما يمكن حساب كل من القيم السابقة باستخدام الدوال الهاشية $f_1(x)$ و $f_2(y)$.

Y \ X	-4	-2	7	**$f_1(x)$**
1	1/8	1/4	1/8	**4/8**
-5	1/4	1/8	1/8	**4/8**
$f_2(y)$	**3/8**	**3/8**	**2/8**	1

$$E(x) = \sum_x x\, f_1(x) = 1(4/8) - 5(4/8) = -2$$

$$V(x) = E(x^2) - E^2(x)$$

$$= [1^2(4/8) + (-5)^2(4/8)] - (-2)^2 = 9$$

3 التباين المشترك Covariance

يعرف التباين المشترك كمايلي:

$$Cov(X, Y) = \sigma_{xy} = E[(X - \mu_x)(Y - \mu_y)]$$

في حالة X و Y متغيرتان متقطعتان:

$$\sigma_{xy} = \sum_x\sum_y (x - \mu_x)(y - \mu_y) f(x, y)$$

في حالة X و Y متغيرتان مستمرتان:

$$\sigma_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - \mu_x)(y - \mu_y) f(x, y)\, dx\, dy$$

أ) خصائص التباين المشترك

1. من تعريف التباين يمكن أن نستنتج:

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

$$= E[XY - X\mu_y - \mu_x Y + \mu_x\mu_y]$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + \mu_x\mu_y$$

$$= E(XY) - E(X)E(Y)$$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

2. في حالة X و Y متغيرتان مستقلتان[1] نعلم من خصائص التوقع الرياضي أن $E(XY) = E(X)$ $E(Y)$ ومنه:

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(XY) - E(XY)$$

$$Cov(X, Y) = 0$$

3. في حالة X و Y متغيرتان مستقلتان أو غير مستقلتين:

$$Var(X \pm Y) = V(X) + V(Y) \pm 2\,Cov(X, Y)$$

4. القيمة المطلقة للتباين المشرك لا تكون أكبر من جداء الانحرافين المعياريين: $|\sigma_{xy}| \le \sigma_x\sigma_y$

5. في حالة X و Y متغيرتان مرتبطتان تماما مثلا $Y = X$ فإن: $Cov(X, Y) = V(X)\,V(Y)$

4 معامل الارتباط

من الخاصية (2) نستنتج أن الكسر $\dfrac{\sigma_{xy}}{\sigma_x\sigma_y}$ يساوي 0 في حالة X وY مستقلتان و من الخاصية (5) نستنتج أنه في حالة X و Y متغيرتان مرتبطتان تماما فإن الكسر يساوي 1.

[1] العكس ليس بالضرورة دوما صحيح، فقد يكون التباين المشترك مساويا للصفر من غير أن يكون المتغيرتان مستقلتان فعلا. المعادلة هي في الحقيقة تمثل شرطا ولكنه ليس شرطا كافيا. بالمقابل يمكن استعمال نتيجة معدومية التباين المشترك للدلالة على ضعف الارتباط، إذاكان موجودا، بين المتغيرتين. الشرط الازم والكافي لاستقلال متغيرتين هو المذكور سابقا في تعريف الاستقلال.

من جهة أخرى من الخاصية 4 نستنتج أن النسبة تتراوح قيمته بين (-1) و(1): $-1 \le \dfrac{\sigma_{xy}}{\sigma_x \sigma_y} \le 1$. من أجل هذا

تستعمل النسبة: $r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$ لقياس الارتباط بين المتغيرتين، وتسمى معامل الارتباط.

في حالة r معدوم نقول أن **المتغيرتان غير مرتبطتين**، من غير أن نجزم أنهما مستقلتان.

مثال. أوجد التباين المشترك والارتباط للتوزيع المشترك المذكور في المثال السابق.

Cov(X, Y) = E(XY) – E(X) E(Y)

E(XY) = 1(-4)(1/8) + (1)(-2)(2/8) + (1)(7)(1/8) + (-5)(-4)(2/8) + (-5)(-2)(1/8) + (-5)(7)(1/8) = 1.75

E(X) = 1(4/8) + (-5)(4/8) = -2,   E(Y) = -4(3/8) – 2(3/8) + 7(2/8) = -1/2.

Cov(X, Y) = 1.75 – (-2)(-1/2) = 0.74

$V(X) = E(X^2) – E^2(X) = 1(4/8) + (-5)^2(4/8) – (-2)^2 = 9 \Rightarrow \sigma_x = 3,$

$V(Y) = E(Y^2) – E^2(Y) = 20.34 \Rightarrow \sigma_y = 4.5.$ $r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y} = \dfrac{0.75}{3(4.5)} = 0.05$

5 خلاصة

نقول عن متغيرتان أنهما مستقلتان إذا أمكن كتابة دالة التوزيع المشتركة لهما (أو دالة الكثافة المشتركة) في شكل جداء دالتين هامشيتين تراكميتين (أو دالتين هامشيتين للكثافة): $f(x, y) = f_1(x)\, f_2(y)$ أي أن:

P(X = x, Y = y) = P(X = x) P(Y = y)

من جهة أخرى نقيس الارتباط بين متغيرتين من خلال معامل الارتباط،

وفي هذه الحالة يكون التباين المشترك Cov(X, Y) = E(XY) – E(X) E(Y)

معدوما لأنه تبعا لخصائص التوقع الرياضي في حالة الاستقلال فإن E(X) E(Y) = E(XY)

لكن العكس ليس بالضرورة صحيح.

يستخدم معامل الارتباط $r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$ كمؤشر على الارتباط بين المتغيرتين، لكن الإحصائي يجب أن يكون متنبها إلى محدودية هذا المؤشر.

الفصل VI. دوال المتغيرات العشوائية والتقارب

الدوال غير الخطية : ك 2، فيشر وستيودنت
التقارب والسلوك التقاربي، نظرية النهاية المركزية

نتناول في المبحث الأول من هذا الفصل عدد من التوزيعات ذات الاستخدام الواسع في الإحصاء الاستدلالي والتطبيقي خاصة في مجالي "التقدير" و"اختبار الفروض"، حيث يستخدم الإحصائي هذه التوزيعات في عمله من أجل الوصول إلى "قرار" بشأن "المجتمع" المدروس انطلاقا من بيانات يتحصل عليها من عينة. في المبحث الثاني سنتطرق للتقارب بين التوزيعات الاحتمالية المختلفة التي درسناها من قبل.

المبحث 1. الدوال غير الخطية: ك 2 ، فيشر وستيودنت

توزيع ك2 توزيع ستيودنت توزيع فيشر

1 توزيع ك2[1] Distribution en Khi-carré (ou Khi-deux)

توزيع ك2 هو من أكثر التوزيعات استخداما في مجال اختبار الفروض بأنواعها، ويمكن تعريفه كما يلي:

لتكن $X_1, X_2, \ldots X_\nu$ ، متغيرات عشوائية مستقلة كل منها تتبع التوزيع الطبيعي المعياري ($\mu = 0, \sigma = 1$). المتغيرة

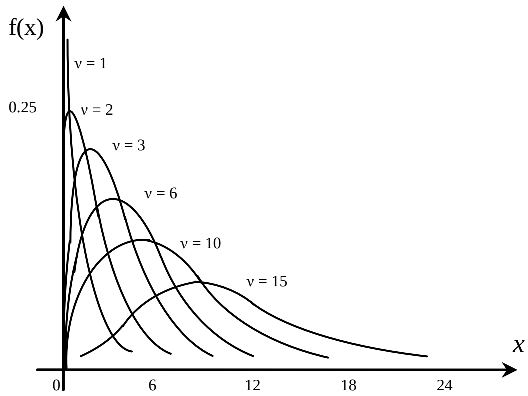
$$X = X_1^2 + X_2^2 + \ldots + X_\nu^2$$

لها دالة الكثافة التالية:

$$\boxed{f(x) = \begin{cases} \dfrac{x^{(\nu/2)-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} & x > 0 \\ 0 & x \le 0 \end{cases}}$$

حيث $\Gamma(\alpha)$ هي الدالة قاما:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\,dt \qquad \alpha > 0$$



رسم 14 تدرج منحنى ك2 حسب درجة الحرية

و نقول أن X تتبع التوزيع ك2 ب ν درجة حرية ونكتب $X \sim \chi^2_\nu$.

الدالة التجميعية $F(\chi^2)$ تكتب كما يلي:

[1] يرجع الفضل في اكتشاف هذا التوزيع إلى ف هلمرت (F. Helmert, 1876) وكارل بيرسون (Karl Pearson, 1900) .

$$P(X^2 \le x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)}\displaystyle\int_0^x u^{(\nu/2)-1}e^{-u/2}du & x \ge 0 \\ 0 & x < 0 \end{cases}$$

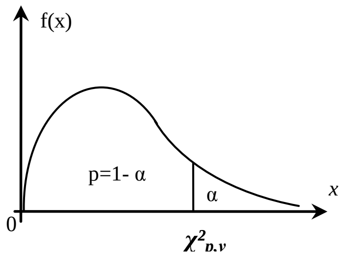
(أ) **خصائص توزيع ك 2**

$E(X) = \nu, \quad V(X) = 2\nu, \quad M(t) = (1-2t)^{-\nu/2}$

دالة التوزيع ك2 هي هي حالة خاصة من توزيع قاما بوضع $\alpha = \nu/2, \beta = 2$.

ويأخذ منحنى $f(x)$ شكله حسب قيمة الثابت ν ونلاحظ من الرسم أن المنحنى يبتعد شيئا فشيئا عن المحور العمودي ويأخذ شكلا جرسيا كلما زادت قيمة ν . ونبرهن أنه عند ν كبير $(\nu \ge 30)$ فإن $\sqrt{2\chi^2} - \sqrt{2\nu - 1}$ تتبع التوزيع الطبيعي المعياري.

- في الجداول الاحصائية، تعين نقطة (قيمة المتغيرة) ك2 على المحور الأفقي (أنظر الرسم المقابل) من خلال ν بالإضافة إلى المساحة p على يسار ك2 تحت المنحنى $(p = P(X \le \chi^2_{\nu;p})$. وأحيانا تحدد النقطة ك2 بدلالة المساحة على يمينها $(\alpha = 1- p)$ لذلك نجد في كتب الاحصاء كل من الكتابتين: $\chi^2_{p,\nu}$ و$\chi^2_{\alpha,\nu}$



رسم 15 تعيين نقطة ك2 على المحور من خلال قيمة p

- **نظرية**: لتكن م ع مستقلة عددها n حيث $X_1 \sim \chi^2_{\nu_1}, \ldots, X_n \sim \chi^2_{\nu_n}$ مجموع هذه المتغيرات

$$X_T = \sum_{i=1}^{n} X_i \sim \chi^2_{\nu_T}$$
$$\nu_T = \sum \nu_i$$

2 توزيع ستيودنت Distribution de Student[1]

لتكن المتغيرتان العشوائيتان المستقلتان Y وZ حيث $Y \sim N(0, 1)$ و $Z \sim \chi^2_\nu$ ؛ المتغيرة $T = \dfrac{Y}{\sqrt{Z/\nu}}$ تتبع توزيع لها دالة الكثافة التالية

[1] يرجع الفضل في إيجاد هذا القانون إلى ويليام سيلي قوسي (1876-1937) (William Sealy Gosset) الذي نشر مقالاته كلها باسم ستيودنت، ونشر مقالته حول هذا القانون عام 1908 بعنوان « The probable error of a mean » أنظر : دراوزبيك 1997، ص 262.

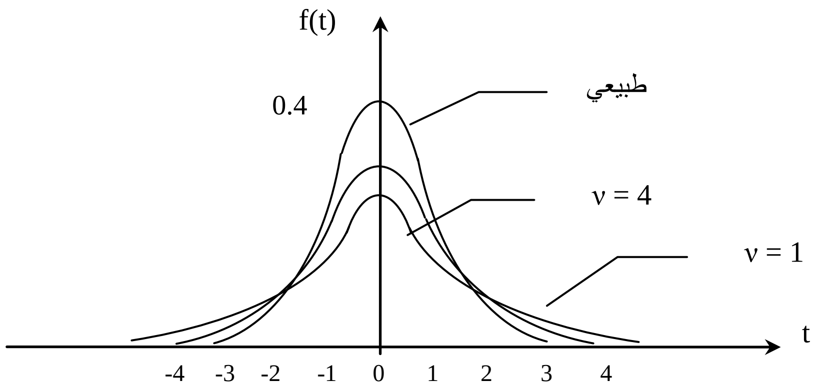
$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\ \Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \qquad -\infty < t < \infty$$

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t}\, dt \qquad \alpha > 0$$

و نقول أن المتغيرة X تتبع توزيع ستيودنت ب ν درجة حرية ونكتب: $T \sim t_\nu$

(أ) **خصائص توزيع ستيودنت**

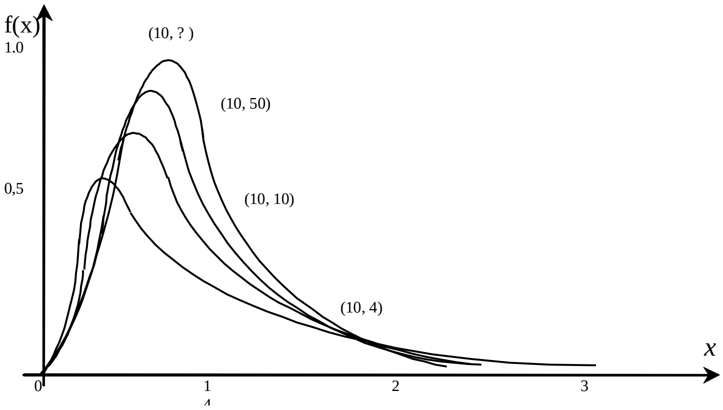
$E(T) = 0, \quad V(T) = \nu/(\nu-2) \quad si \quad (\nu > 2)$



رسم16 تدرج منحنى ستيودنت حسب درجة الحرية

- نلاحظ أن منحنى t متماثل حول المتوسط 0 مما يعني أن لكل نقطة موجبة t نقطة مناظرة لها سالبة حيث المساحة تحت المنحنى على يمين t تساوي المساحة تحت المنحنى علي يسار (–t)، ونكتب $t_{1-p} = - t_p$.
- بالإضافة إلى ذلك فإن منحنى f(t) يقترب من المنحنى الطبيعي المعياري كلما زادت قيمة ν . وعموما، يعتبر الإحصائيون أن المنحنيان يتطابقان تقريبا عند $\nu \geq 30$.
- في الجداول الاحصائية، تعين نقطة (قيمة المتغيرة) t من خلال ν والمساحة p على يسار t تحت المنحنى ($(p = P(T \leq t_{\nu;p})$. وأحيانا تحدد النقطة t بدلالة المساحة على يمينها $(\alpha = 1 - p)$ ونكتب : $t_{p,\nu}$ أو $t_{\alpha,\nu}$.

3 توزيع فيشر Distribution F de Fisher-Snédecor[1](F)



رسم 17 تدرج منحنى فيشر حسب درجة الحرية

ليكن لدينا المتغيرتان العشوائيتان المستقلتان $X_1 \sim \chi^2_{\nu_1}$ و $X_2 \sim \chi^2_{\nu_2}$. المتغيرة : $X = \dfrac{X_1/\nu_1}{X_2/\nu_2}$ لها دالة الكثافة:

$$\boxed{f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)} \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} x^{(\frac{\nu_1}{2})-1} (\nu_2+\nu_1 x)^{-\frac{\nu_1+\nu_2}{2}} & x>0 \\ 0 & x \le 0 \end{cases}}$$

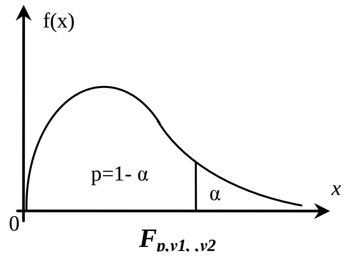
و نقول أن المتغيرة X تتبع توزيع فيشر ب ν_1 وν_2 درجة حرية ونكتب: $X \sim F_{\nu 1,\, \nu 2}$

(أ) **خصائص توزيع فيشر**:

$$\mu = \frac{\nu_2}{\nu_2-2} \quad (\nu_2>2) \quad , \quad \sigma^2 = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)(\nu_2-2)^2} \quad (\nu_2>4)$$

ويظهر من المعادلة تبعية منحنى f(x) بالإضافة ل x إلى كل من ν_1 وν_2 ولذلك تحدد أي نقطة F من خلال ثلاثة معالم: ν_1 وν_2, p (المساحة تحت المنحنى على يسار النقطة F) ، ونكتب F_{p,ν_1,ν_2}

وفي الغالب تعطي الجداول الإحصائية قيم F عند p = 0.95 وp = 0.99 .



رسم18 تعيين قيمة F يتم في الجدول يتم من خلال ν_1, ν_2 و P

[1] رونالد آيلمر فيشر (1890-1962) (Ronald Aylmer Fisher) (إنجلترا) يعتبر مؤسس نظرية التقدير و جورج وادل سنيديكور (George Waddel Snédecor) (1881-1974) (أمريكي) أنظر المرجع السابق، ص 258.

نظرية 1.

$$F_{1-p,\nu_1,\nu_2} = 1/F_{p,\nu_2,\nu_1}$$

نظرية 2.

$$F_{1-p,\,1,\,\nu} = t^2_{1-(p/2),\nu}$$

نظرية 3.

$$F_{p,\nu,\,\infty} = \frac{\chi^2_{p,\nu}}{\nu}$$

4 خلاصة

يمكن تلخيص أهم ما تضمنه هذا المبحث في الجدول التالي:

التوزيع	المتغيرة العشوائية	أهم ما يجب معرفته عن دالة الكثافة
توزيع ك2 $X \sim \chi^2_\nu$	إذا كانت Xi متغيرات عشوائية مستقلة كل منها تتبع التوزيع الطبيعي المعياري، و $X = X_1^2 + X_2^2 + \ldots + X_\nu^2$ إذن: $X \sim \chi^2_\nu$	$f(x) = 0 \quad si \quad x \leq 0$ $E(X) = \nu, \quad V(X) = 2\nu$
توزيع ستيودنت $T \sim t_\nu$	لتكن المتغيرتان العشوائيتان المستقلتان Y وZ حيث $Y \sim N(0, 1)$ و $Z \sim \chi_\nu^2$ ؛ إذن: $T = \frac{Y}{\sqrt{Z/\nu}} \sim t_\nu$	$E(T) = 0,$ $V(T) = \nu/(\nu-2) \quad si \ (\nu > 2)$
توزيع فيشر $X \sim F_{\nu1,\,\nu2}$	إذا كانت لدينا متغيرتان عشوائيتان مستقلتان حيث: $X_1 \sim \chi_{\nu_1}^2$ و $X_2 \sim \chi_{\nu_2}^2$ ، فإن $X = \frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu1,\,\nu2}$	$f(x) = 0 \quad si \quad x \leq 0$ $F_{p,\nu,\,\infty} = \frac{\chi^2_{p,\nu}}{\nu}$, $F_{1-p,\,1,\,\nu} = t^2_{1-(p/2),\nu}$; $F_{1-p,\nu_1,\nu_2} = 1/F_{p,\nu_2,\nu_1}$

المبحث 2. السلوك التقاربي لبعض التوزيعات الاحتمالية

التقارب بين التوزيع الثنائي والتوزيع الطبيعي
الانتقال من متغيرة متقطعة إلى متغيرة مستمرة
التقارب بين التوزيع الثنائي وتوزيع بواسون
نظرية النهاية المركزية

نتناول في هذا المبحث بعض حالات التقارب الذي يحصل بين عدد من التوزيعات الاحتمالية الشهيرة. ونقصد بالتقارب بين توزيعين (الثنائي وبواسون مثلا) أن يعطي التوزيعان نتائج متقاربة بخصوص احتمال معين، مما يعني إمكانية استخدام

توزيعين احتماليين (وأحيانا أكثر) لحساب احتمال معين. علما أننا قد تطرقنا من قبل بإيجاز إلى هذا المفهوم عند دراستنا لهذه التوزيعات.

1 التقارب بين التوزيع الثنائي والتوزيع الطبيعي

لندرس السلوك التقاربي لمتغيرة التوزيع الثنائي X~B(n,p) عندما تؤول n إلأعداد كبيرة جدا.

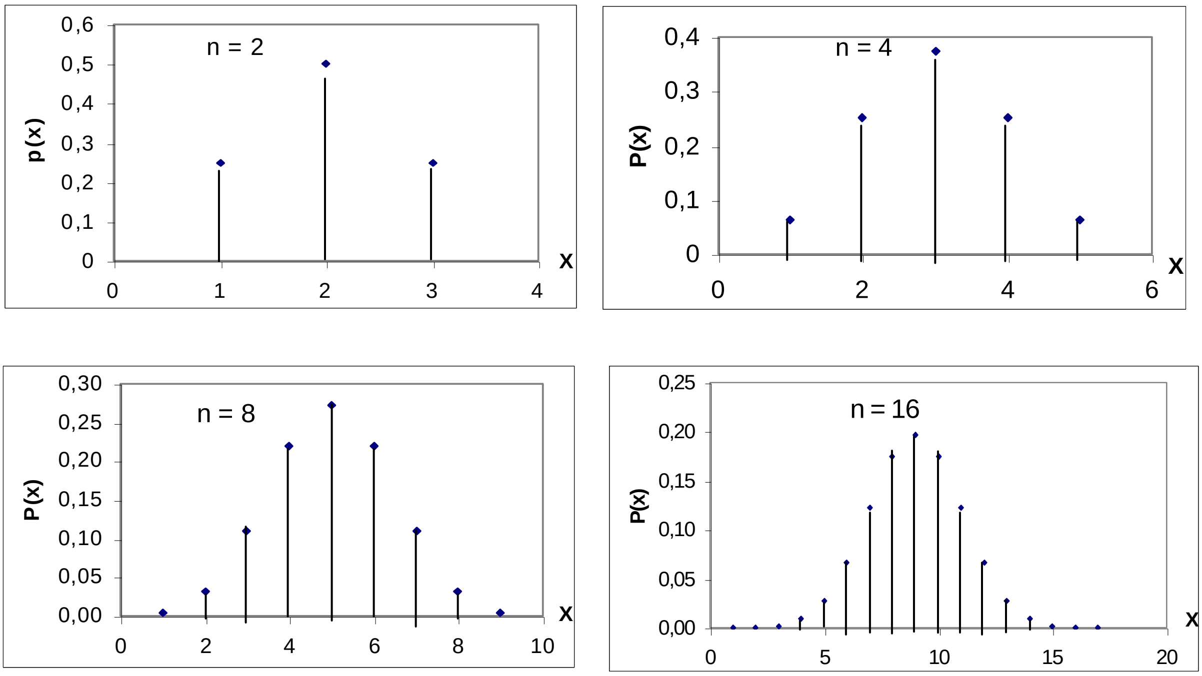
ليكن X يمثل عدد مرات الحصول على صورة عند رمي قطعة نقدية : مرتين، 4 مرات، 8 مرات، 16 مرات.

Xi	0	1	2
P_i	1/4	1/2	¼

Xi	0	1	2	3	4
P_i	1/16	4/16	6/4	4/16	1/16

Xi	0	1	2	3	4	5	6	7	8
P_i	0,004	0,031	0,109	0,219	0,273	0,219	0,109	0,031	0,004

برسم منحنيات Pi للحالات n = 2 ، n = 4 ، n = 8 ، n =16 يظهر السلوك التقاربي للمتغيرة X .



رسم 19 السلوك التقاربي للتوزيع الثنائي لما p = 0.5

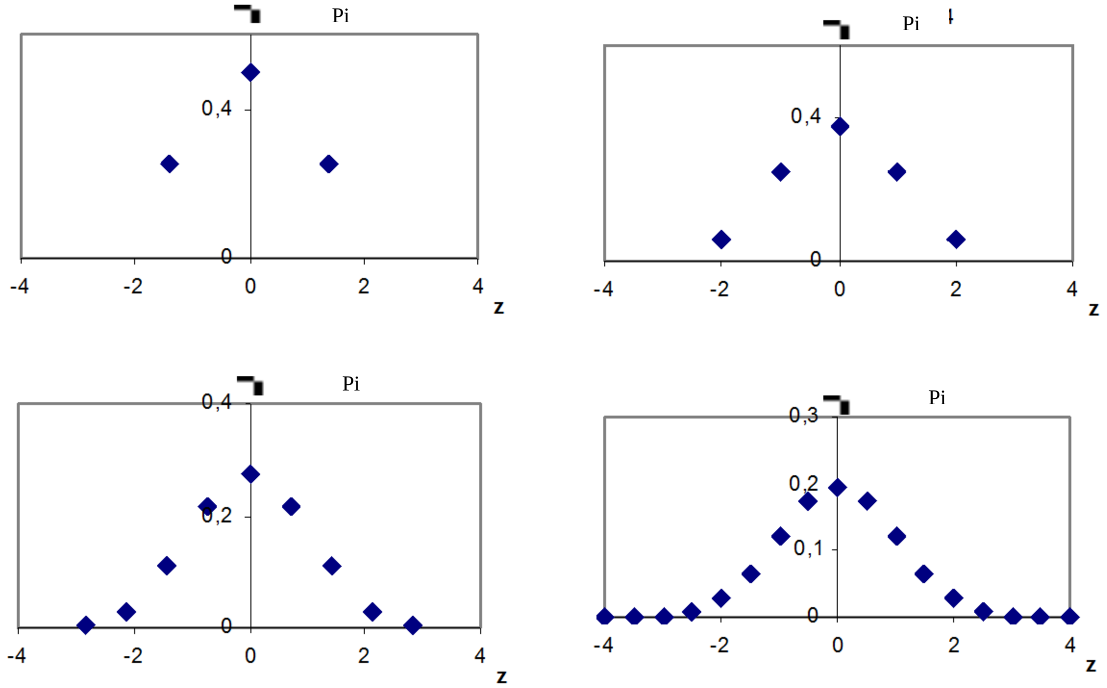
يظهر من مقارنة المنحنيات الأربعة أن زيادة قيمة n تؤدي إلى الحصول على منحنى ذا شكل جرسي ومتماثل حول التوقع μ .

هذه الملاحظة تصدق أيضا في حالة $p \neq 0.5$ لكن التحول يكون أكثر بطأ.

من أجل التعميم نعتبر المتغيرة المعيارية $z = (x - \mu)/\sigma$ الملحقة بذات المتغيرة ذات التوزيع الثنائي X . إن السلوك التقاربي لـZ الملاحظ في الشكل أسفله هو ما تثبته النظرية التالية:

$$soit\ \ X \sim \mathbf{B}(n,p): \qquad Y = \frac{X - np}{\sqrt{npq}} \xrightarrow[n \longrightarrow \infty]{} N(0,1).$$

ونكتب $Y \approx N(0,1)$.



رسم 20 السلوك التقاربي للتوزيع الثنائي من خلال المتغيرة المعيارية المعيارية

قاعدة:

في حالة n كبيرة و p غير قريب من 0 يمكن اعتبار التوزيع الثنائي كتقريب جيد للتوزيع الطبيعي. ويعطي التوزيعان نتائج أكثر تقاربا كلما كانت n كبيرة أكثر. ونكتب:

$$\lim_{n \longrightarrow \infty} P(a \leq \frac{x-np}{\sqrt{npq}} \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-u^2}{2}} du$$

و مما يسرع تقارب التوزيع الثنائي من التوزيع الطبيعي كون p قريب من 0.5 وكقاعدة :

- **عموما نعتبر أن التقريب ملائم عندما np و nq كلاهما أكبر من 5.**
- عدد من الاحصائيين[1] يعتمد قاعدة أخرى هي أن يكون أحد الشرطين التاليين متوفرين:
  - $npq \geq 9$
  - $n \geq 20$, $np \geq 10$, $nq \geq 10$

في حالة $p = 0.5$ ، الشرط (1) يتحقق عند $n = 36$ والثاني عند $n = 20$.

في حالة $p = 0.10$ ، الشرطين يتحققان عند $n = 100$.

2 الانتقال من متغيرة متقطعة إلى متغيرة متصلة.

لاستخدام التوزيع الطبيعي بدلا من التوزيع الثنائي يعني حساب الاحتمال عن طريق توزيع مستمر بينما المتغيرة متقطعة. من أجل ذلك يتم اعتبار كل قيمة في المتغيرة الأصلية مجالا.

[1] المرجع السابق، ص 262.

مثال. احتمال 4 نجاحات خلال n تجربة يصاغ كما يلي: $P(3.5 \le X \le 4.5)$.

مثال2: نرمي قطعة نقدية 20 مرة. ليكن X عدد مرات الحصول على صورة. أحسب $P(X = 8)$ ثم أدرس إمكانية استخدام نظرية موافر- لابلاس لحساب نفس الاحتمال.

$$P(X = 8) = F(8) - F(7) = 0.2517 - 0.1316 = 0.1201. \quad , X \sim B(20, 0.5)$$

لدينا $np = 10 > 5$ وكذلك $nq = 10 > 5$ ، وإذا شئنا استخدام القاعدة الثانية فإننا نجد أيضا أن : $n = 10$، $np = 10$، $nq = 10$، يمكن إذا اعتبار $Y = (X-10)/\sqrt{5} \sim N(0, 1)$. نستخدم المتغيرة المستمرة X^* بدلا من X لحساب احتمال المجال المعبر عن القيمة 8 وهو [7.5, 8.5]

$$P(7.5 \le X^* \le 8.5) = P\left(\frac{7.5-10}{2.24} \le Z \le \frac{8.5-10}{2.24}\right) = P(-1.12 \le Z \le -6.67) = 0.12$$

3 التقارب بين التوزيع الثنائي وتوزيع بواسون

يعطي توزيع بواسون نتائج قريبة من التوزيع الثنائي لما $n \ge 30$ و $np < 5$ أو $nq < 5$

و يستخدم بعض الإحصائيين كشرط لاستعمال قانون بواسون بدلا من القانون الثنائي القاعدة التالية[1]:

$n \ge 25$ و $p \le 0,1$

مثال : 10 % من إنتاج آلة ما يعد تالفا، نأخذ 30 وحدة من انتاج هذه الآلة عشوائيا.

أحسب احتمال أن يكون هناك وحدتان تالفتان.

$$P(X = 2) = C^2_{30}\,(0,1^2)\,(0.9^{28}) = 0.22$$

لدينا $n \ge 25$، $p \le 0.1$: لاستعمال توزيع بواسون نحسب أولا قيمة المعلمة (معلمة قانون بواسون)

$$\lambda = \mu = np = 30 * 0,1 = 3$$

$$P(2) = \lambda^x * e^{-\lambda}/x! = (3^2 * e^{-3}) / 2! = 0.22$$

4 نظرية النهاية المركزية

لتكن المتغيرات X_1، X_2، . . . . متغيرات عشوائية مستقلة لها نفس التوزيع الاحتمالي بتباين ومتوسط محددين. إذا كانت

$$S_n = X_1 + X_2 + \ldots + X_n \qquad (n = 1, 2, \ldots),$$

فإن S_n تتبع التوزيع الطبيعي عندما $n \to \infty$. وبما أن $E(S_n) = n\mu$ و $\sigma_{S_n} = \sigma\sqrt{n}$ فإننا نكتب :

$$\lim_{n \longrightarrow \infty} P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{\frac{-z^2}{2}}\,dz$$

في الحقيقة فإن النظرية محققة عندما تكون المتغيرات المستقلة Xi لها نفس المتوسط والتباين حتى ولو لم يكن لها بالضرورة نفس التوزيع، مع العلم أنه توجد صيغ أخرى لهذه النظرية حيث لا يشترط أن يكون للمتغيرات نفس التوزيع الاحتمالي ولا حتى أن تكون مستقلة.

تجدر الإشارة إلى أن نظرية موافر- لابلاس التي تطرقنا إليها سابقا هي حالة خاصة من نظرية النهاية المركزية، ذلك أن متغيرة تتبع القانون $B(n, p)$ يمكن اعتبارها مجموعا لعدد من المتغيرات المستقلة ذات التوزيع البرنولي $B(1, p)$.

---

[1] المرجع السابق

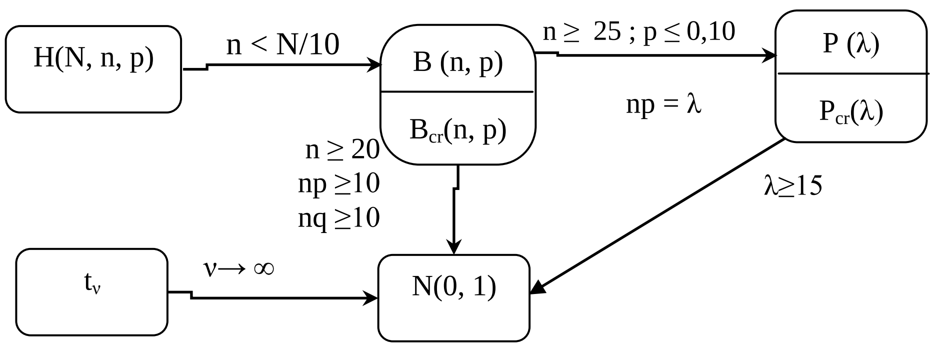
5 خلاصة

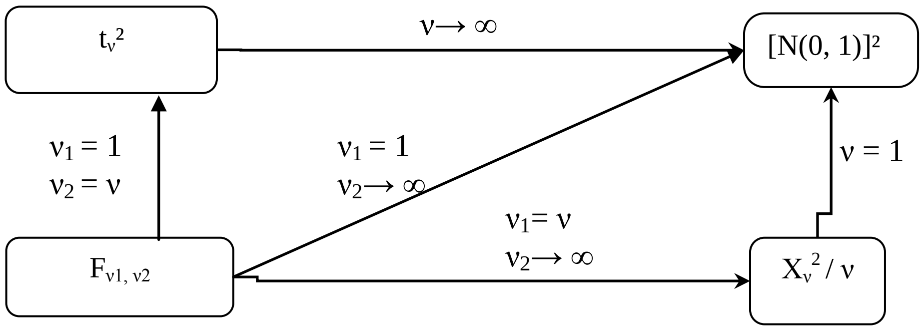
لاستخدام التوزيع الطبيعي بدلا من التوزيع الثنائي يعني حساب الاحتمال عن طريق توزيع مستمر بينما المتغيرة متقطعة. من أجل ذلك يتم اعتبار كل قيمة في المتغيرة الأصلية مجالا.

نظرية النهاية المركزية تنص على أن S_n (متغيرات عشوائية مستقلة لها نفس التوزيع الاحتمالي بتباين ومتوسط محددين) تتبع التوزيع الطبيعي عندما $n \to \infty$ بمتوسط $E(S_n) = n\mu$ و $\sigma_{S_n} = \sigma\sqrt{n}$ ونكتب:

$$\lim_{n \longrightarrow \infty} P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-z^2}{2}} dz$$

الرسم البياني التالي يبين القواعد المستخدمة كشروط للتقريب بين التوزيعات الاحتمالية المذكورة آنفا في المبحث بالإضافة إلى التوزيعات الأخرى التي درست في الفصول السابقة (الرمز cr يعني متغيرة معيارية).





رسم يبين قواعد التقريب بين القوانين الاحتمالية الأكثر

الفصل VII. نظرية توزيع المعاينة

مفاهيم إحصائية
توزيعات المعاينة للمتوسطات
توزيع المعينة للنسبة
توزيع المعاينة للفروق و المجاميع
توزيع المعاينة للتباين و توزيع المعاينة لنسبة تباينين

تنتشر في مجتمعاتنا المعاصرة عمليات الاستقصاء، ففي عالم الأعمال تقوم المؤسسات عن طريق مصالح التسويق ومصالح البحوث والتطوير بإجراء استقصاءات للإطلاع على توجهات المستهلكين، وفي وسائل الإعلام لا يمر يوم دون أن يعلن عن نتائج استقصاء أجرته مجلة أو جامعة حول مواضيع سياسية أو اجتماعية متعددة، منها الاستقصاءات المثيرة للجدل حول الأراء السياسية للمواطنين أثناء الحملات الانتخابية. فما هي الأسس النظرية الرياضية التي تستند عليها الاستقصاءات المختلفة ؟ أو كيف يمكن الاستدلال من خلال بيانات عينة على خصائص المجتمع الذي أخذت منه؟ الإجابة على هذه الأسئلة و غيرها تتطلب فهم العلاقات الرياضية بين الخصائص المختلفة للمجتمع مثل المتوسط، التباين وغيرها، والخصائص المناظرة لها في العينة وهو ما سنتناوله في هذا الفصل. في الفصول المقبلة سندرس عددا من التطبيقات لهذه العلاقات الرياضية.

المبحث 1. مفاهيم إحصائية

المجتمع والعينة
العينة النفادية والعينة غير النفادية
العينة العشوائية
معالم مجتمع
إحصائية المعاينة

1 المجتمع والعينة Population et échantillon

نشرح هذين المصطلحين من خلال الأمثلة التالية:

- قد ترغب الإدارة العسكرية في تقدير الوزن المتوسط للجندي، فتقوم أخذ أوزان عينة من 100 جندي من بين مجموع الجنود (المجتمع).
- ترغب هيأة معينة بالبحوث السياسية في تقدير نسبة الناخبين المساندين لمرشح معين في 10 الولايات، فتقوم باستجواب 100 ناخب من كل ولاية. الناخبون في الولايات العشر يمثلون المجتمع بينما ال 1000 ناخب المستجوبون يمثلون العينة.
- من أجل معرفة مدى دقة صنع قطعة نقدية ترمى القطعة 100 مرة ونحسب عدد مرات الحصول على الصورة والكتابة، حجم العينة هنا هو 100.
- لتقدير نسبة الكرات داخل صندوق، التي من لون معين، نقوم عدد من المرات بسحب كرة نسجل لونها ثم نعيدها. عدد الكرات المسحوبة يمثل حجم العينة.

نلاحظ أن مصطلح المجتمع يقصد به القياسات أو القيم وليس الأفراد أو الأشياء التي تم قياسها (مجتمع الأوزان، مجتمع آراء الناخبين..)، كما أن المجتمع قد يكون محدودا أو غير محدود (نتائج رميات قطعة النقد)، أما العينة فهي عادة تكون محدودة، ونرمز عادة لحجم المجتمع ب N، ولحجم العينة ب n.

2 العينة النفادية والعينة غير النفادية Echantillon exhaustif et non exhaustif

عندما يكون السحب بالإرجاع حيث يمكن أن تظهر المفردة أكثر من مرة في العينة، نسمي هذه المعاينة غير نفادية لأن تكرار العملية لا يؤدي إلى تقليص عدد المفردات في المجتمع، والعكس نسمي المعاينة بدون إرجاع معاينة نفادية. هناك فرضيتان تتكرران في عدد من العلاقات الرياضية التي سنراها لاحقا، هما فرضية أن قيم مفردات العينة مستقلة والمجتمع لانهائي. يتحقق شرط الاستقلال إذا كانت المعاينة غير نفادية، وإذا كانت كذلك، يمكن اعتبار المجتمع مجتمعا غير محدود.

3 العينة العشوائية Echantillon aléatoire

من أجل أن تكون العينة ممثلة للمجتمع، أحد الطرق المستخدمة هي العينة العشوائية. نظريا (قد يصعب تحقيق ذلك في الواقع)، نقول عن عينة أنها عشوائية إذا كان لكل مفردة في المجتمع نفس الاحتمال لأن تكون في العينة. تسمى هذه العينة بالعينة العشوائية البسيطة. لإنجاز ذلك إما أن نسحب المفردات بطريقة عشوائية أو نرقم مفردات المجتمع ثم نحدد العينة من خلال مجموعة من الأعداد تؤخذ من الجداول الإحصائية للأعداد العشوائية[1].

4 معالم المجتمع Paramètre d'une population

نقصد بمعالم المجتمع مجموعة من خصائصه مثل المتوسط، التباين، معامل التماثل، ... من خصائص المجتمع أيضا طبيعة توزيعه الاحتمالي $f(x)$ كأن يكون طبيعيا أو غيره.

5 إحصائية المعاينة Statistique de l'échantillonnage

لتقدير معالم المجتمع (متوسط المجتمع μ، تباين المجتمع σ^2 النسبة p ...) ننطلق من بيانات العينة، حيث نحتاج إلى حساب معالم مثل متوسط العينة m ، تباين العينة S^2، النسبة في العينة p'. بصفة عامة، نسمي كل قيمة تحسب انطلاقا من بيانات العينة من أجل تقدير قيمة معالم المجتمع إحصائية المعاينة. نظريا (رياضيا) إحصائية المعاينة هي كل دالة في المتغيرات العشوائية التي تمثل القيم المحصل عليها في العينة.

المبحث 2. توزيع المعاينة للمتوسطات

متوسط توزيع المعاينة للمتوسطات
تباين توزيع المعاينة للمتوسطات
طبيعة توزيع المعاينة للمتوسطات

1 متوسط توزيع المعاينة للمتوسطات

مسألة: ليكن المجتمع 1، 3، 5، 6، 8. ما هي القيمة المتوقعة لمتوسط عينة مسحوبة بالإرجاع مكونة من مفردتين (m)؟ أحسب متوسط المجتمع μ. قارن بين m وμ. من أجل تحديد ذلك أحسب جميع الحالات الممكنة للمتوسط m_i حسب كل عينة.

العينات الممكنة العينات الممكنة ذات الحجم $n = 2$ من مجتمع حجمه 5 عددها: $5*5 = 25$

[1] أنظر جدول الأعدلد العشوائية.

المتوسطات الممكنة للعينة (معاينة غ نفادية) m_i				
1	2	3	3,5	4,5
2	3	4	4,5	5,5
3	4	5	5,5	6,5
3,5	4,5	5,5	6	7
4,5	5,5	6,5	7	8

العينات الممكنة				
(1, 1)	(3, 1)	(5, 1)	(6, 1)	(8, 1)
(1, 3)	(3, 3)	(5, 3)	(6, 3)	(8, 3)
(1, 5)	(3, 5)	(5, 5)	(6, 5)	(8, 5)
(1, 6)	(3, 6)	(5, 6)	(6, 6)	(8, 6)
(1, 8)	(3, 8)	(5, 8)	(6, 8)	(8, 8)

القيمة المتوقعة m ل m_i هي متوسط قيمها وهي $m = (\sum_i m_i) / 25 = 4,6$.

حساب متوسط المجتمع: $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

مثال2. أوجد نفس مطالب المثال 1. في حالة السحب بدون إرجاع. العينات الممكنة عددها: $C^2_5 = 10$

المتوسطات الممكنة للعينة أو **توزيع المعاينة للمتوسطات (معاينة نفادية)** m_i			
2			
3	4		
3,5	4,5	5,5	
4,5	5,5	6,5	7

العينات الممكنة بدون إرجاع			
(1, 3)			
(1, 5)	(3, 5)		
(1, 6)	(3, 6)	(5, 6)	
(1, 8)	(3, 5)	(5, 8)	(6, 8)

القيمة المتوقعة m ل m_i هي متوسط قيمها وهي:

$$E(m) = \mu_m = (\sum_i m_i) / 10 = 4,6$$

متوسط المجتمع : $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

نظرية 1. إذاكانت م ع تمثل مجتمع ما وm متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لمتوسط العينة E(M) تكتب كما يلي: $E(M) = \mu_m = \mu$

البرهان : لنرمز ب Xi لقيم المتغيرة الأصلية X .

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n}\sum_i E(Xi) = \frac{1}{n}\sum_i \mu = \frac{1}{n}n\mu = \mu.$$

2 تباين توزيع المعاينة للمتوسطات

(أ) حالة المعاينة بالإرجاع

مثال. أحسب تباين المجتمع في المسألة 1، أحسب التباين (والانحراف المعياري) لتوزيع المعاينة للمتوسطات σ^2_m علما أن العينة مسحوبة بالإرجاع (غ نفادية)، قارن بين تباين المجتمع وتباين متوسطات العينات الممكنة (توزيع المعاينة للمتوسطات).

mi				
1	2	3	3,5	4,5
2	3	4	4,5	5,5
3	4	5	5,5	6,5
3,5	4,5	5,5	6	7
4,5	5,5	6,5	7	8

$\sigma^2_m = [\sum_i (m_i - m)^2]/25 = 2.92;$

$\sigma^2 = [\sum_i (x_i - \mu)^2]/5 = 5.84$

2.92 = 5.84 / 2

هذا المثال يمهد للنظرية التالية:

نظرية 2. إذا كانت م ع تمثل مجتمع ما وm_i متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع بالإرجاع، فإن تباين m_i (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي: $\sigma_m^2 = \frac{\sigma^2}{n}$ حيث n حجم العينة.

البرهان: لنرمز ب Xi لقيم المتغيرة الأصلية X .

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n^2}\sum_i V(Xi) = \frac{1}{n^2}\sum_i \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

(ب) **حالة المعاينة بدون إرجاع.**

مسألة: في المسألة 1 أحسب تباين المتوسطات الممكنة للعينة σ^2_m في حالة المعاينة بدون إرجاع، قارن بين تباين المجتمع وتباين المتوسطات الممكنة للعينة.

المتوسطات الممكنة للعينة أو توزيع المعاينة للمتوسطات (معاينة نفادية) mi			
			2
		4	3
	5,5	4,5	3,5
7	6,5	5,5	4,5

تباين المتوسطات الممكنة للعينة:

$\sigma^2_m = [\sum_i (m_i - m)^2]/10 = 2.19$

تباين المجتمع: $\sigma^2 = [\sum_i (x_i - \mu)^2]/5 = 5.84$

(أو بطريقة ثانية:

$\sigma^2 = E(X^2) - E(X)^2$

$= (1 + 9 + 25 + 36 + 64) / 5 - 4.6^2 = 5.84)$

المقارنة بين تباين متوسط العينة و تباين المجتمع:

$$2.19 = \frac{5.84}{2}\left(\frac{5-2}{5-1}\right)$$

هذا يمهد للنظرية التالية:

نظرية 3. إذا كانت X م ع تمثل مجتمع ما حجمه N وm_i متغيرة ع تمثل متوسط عينة حجمها n مسحوبة من ذات المجتمع بدون إرجاع، فإن تباين m_i (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي:

$$\sigma_m^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

وتسمى النسبة $\frac{N-n}{N-1}$ معامل الإرجاع.

3 طبيعة توزيع m

ندرس طبيعة توزيع متوسط توزيع المعاينة للمتوسطات من خلال النظريات التالية:

نظرية 4. إذا كان المجتمع موزع طبيعيا بمتوسط μ وتباين σ^2 فإن متوسط العينة المسحوبة منه يتبع أيضا التوزيع الطبيعي بمتوسط μ وتباين σ^2/n، ونكتب $m \approx N(\mu, \sigma^2/n)$

نظرية 5. (**نظرية النهاية المركزية**): إذا كان المجتمع الذي تسحب منه العينة ذو متوسط μ وتباين σ^2 لكن ليس بالضرورة طبيعيا فإن المتغيرة المعيارية ل m أي $z = \dfrac{m-\mu}{\sigma/\sqrt{n}}$ تؤول إلى التوزيع الطبيعي المعياري عندما يكون n كبيرا (n ≥ 30) ونكتب:
$z \approx N(0, 1)$.

في حالة المجتمع محدود والمعاينة نفادية نستبدل العبارة $\sigma/\sqrt{n}$ ب

$$\sigma_m = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

عمليا يستخدم الإحصائيين هذه الصيغة المعدلة بمعامل الإرجاع للانحراف المعياري عندما $n/N \geq 0.05$

مثال: مجتمع حجمه 900 بمتوسط $\mu = 20$ و$\sigma = 12$. نستخرج كل العينات الممكنة. أحسب المتوسط والانحراف المعياري لتوزيع المعاينة للمتوسطات في حالة: (1) حجم العينة $n = 36$، (2) $n = 64$.

(1) $n = 36$: $n/N = 36/900 = 0.04 < 0.05 \Rightarrow \sigma_m = \sigma/\sqrt{n} = 12/\sqrt{36} = 2$

$$(2) \quad n = 64 : N = 900 \Rightarrow \frac{n}{N} = \frac{64}{900} = 0.071 > 0.05$$

$$\Rightarrow \sigma_m = \frac{12}{\sqrt{64}}\sqrt{\frac{900-64}{900-1}} = 1.92$$

$E(m) = \mu = 20$

مثال2. باستخدام معطيات المثال السابق ($n = 36$) أحسب احتمال أن يكون m محصورا بين 18 و22.

أحسب نفس الاحتمال في حالة $n = 64$.

$$Z_1 = \frac{m_1 - \mu}{\sigma/\sqrt{n}} = \frac{18-20}{12/\sqrt{36}} = -1 \;, \quad Z_2 = 1 \;\Rightarrow P(18 < m < 22) = P(Z_1 < Z < Z_2) = 0.6827$$

$$Z_1 = \frac{m_1 - \mu}{\sigma/\sqrt{n}}\left(\frac{N-n}{N-1}\right) = \frac{18-20}{1.92} = -1.04 \;, \quad Z_2 = 1.04 \;\Rightarrow P(18 < m < 22) = P(-1.04 < Z < 1.04) = 0.70$$

4 خلاصة

الجدول التالي يبين أهم خصائص توزيع المعينة للمتوسطات.

المجتمع	المعاينة	الخاصية
مجتمع ما	سحب بالإرجاع أو بدون إرجاع	$E(M) = \mu_m = \mu$
مجتمع ما	سحب بالإرجاع	$\sigma_m^2 = \frac{\sigma^2}{n}$
مجتمع ما حجمه N	سحب بدون إرجاع	$\sigma_m^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$
مجتمع موزع طبيعيا بمتوسط μ وتباين σ²	سحب بالإرجاع أو بدون إرجاع	$m \approx N(\mu, \sigma^2/n)$
مجتمع بمتوسط μ وتباين σ² لكن ليس بالضرورة طبيعيا	عندما يكون n كبيرا $(n \geq 30)$	$z = \frac{m-\mu}{\sigma/\sqrt{n}} \approx N(0, 1)$

المبحث 3. توزيع المعاينة للنسبة

النظرية التالية تبين المتوسط، التباين، و طبيعة التوزيع الإحصائية **p'** : نسبة خاصية ما في العينة.

نظرية 6 : لتكن X م ع تمثل مجتمع ما غير محدود وموزع طبيعيا حيث p نسبة المفردات في المجتمع ذات صفة معينة، ولتكن p' م ع تمثل نسبة المفردات ذات الصفة المذكورة في العينة المسحوبة من ذات المجتمع، نحصل على توزيع للإحصائية p' حيث معالمه E(p') و $\sigma_{p'}$، هذه المعالم تساوي : $\sigma^2{}_{p'} = \frac{pq}{n}$; $E(p') = \mu_{p'} = p$

عند $n \geq 30$: $p' \approx N(p, \sigma_{p'})$

عندما يكون المجتمع محدودا والمعاينة نفاديه نضرب في معامل الإرجاع عند حساب الانحراف المعياري.

مثال[1]. لاحظت إدارة الجامعة أنه في عينة من 100 طالب، 40 حصلوا أخيرا على شهادة. تريد الإدارة تقدير نسبة الطلبة الذين يحصلون على الشهادة داخل مجال يكون احتماله 90 بالمائة.

$$P(p_1 < p' < p_2) = 0.9 \; ; \; n \geq 30,$$

نفترض أن N كبير بحيث : $n/N < 0.05$

$$\Rightarrow p' \sim N(p, \sigma_{p'}),\ \sigma_{p'} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4(0.6)}{100}} \cong 0.05$$

$$P(p_1 < p' < p_2) = 0.9 = P(z_1 < Z < z_2) \Rightarrow z_1 = -1.64, \qquad z_2 = 1.64$$

$$Z1 = \frac{(p_1 - p)}{\sigma_{p'}} \Rightarrow p = p' \pm z\,(\sigma_{p'}) = 0.4 \pm 1.64(0.05) = 0.4 \pm 0.082$$

$$\Rightarrow P(0.318 < p < 0.482) = 0.9.$$

[1] م سبياجال 1985 ص 82.

المبحث 4. توزيع المعاينة للفروق والمجاميع

متوسط و تباين توزيع المعاينة للفروق و المجاميع
طبيعة توزيع المعاينة للفروق و المجاميع

1 المتوسط والتباين

ليكن لدينا مجتمعين نسحب من كل منهما عينة عشوائية، نحسب في كل عينة محسوبة من المجتمع الأول الإحصائية S_1 ونحسب نفس الاحصائية (المتوسط مثلا أو التباين ...) في كل عينة من المجتمع الثاني ونسميها S_2. إن الفرق $S_1 - S_2$ يشكل بدوره متغيرة عشوائية لها المتوسط والتباين التاليين:

$$\mu_{S-S2} = \mu_{S1} - \mu_{S2} \qquad \sigma^2_{S1-S2} = \sigma^2_{S1} + \sigma^2_{S2}$$

مثال 1. إذا كانت الاحصائية هي **المتوسط** فإن:

$$\mu_{m1-m2} = \mu_{m1} - \mu_{m2} = \mu_1 - \mu_2 \qquad \sigma^2_{m1-m2} = \sigma^2_{m1} + \sigma^2_{m2} = \sigma^2/n_1 + \sigma^2/n_2$$

مثال 2. إذا كانت الاحصائية هي **النسبة** فإن:

$$\mu_{p1-p2} = \mu_{p1} - \mu_{p2} = p_1 - p_2 \qquad \sigma^2_{p1-p2} = \sigma^2_{p1} + \sigma^2_{p2} = p_1q_1/n_1 + p_2q_2 / n_2$$

إذا كان الاهتمام هو على مجموع الاحصائيتين بدلا من الفرق بينهما فإن:

$$\mu_{S1+S2} = \mu_{S1} + \mu_{S2} \qquad \sigma^2_{S1+S2} = \sigma^2_{S1} + \sigma^2_{S2}$$

2 طبيعة توزيع المعاينة للفرق بين متوسطين

> **نظرية 7 :** في حالة $n_1 \geq 30$ وn_2 ، يقترب توزيع المتغيرة المعيارية للفرق بين متوسطين من التوزيع الطبيعي المعياري. ونكتب: $\mu_{m1-m2} \approx N(0, 1)$

مثال[1] : ليكن المجتمع U_1 : 3، 7، 8. والمجتمع U_2 : 2، 4. تحقق من أن :

$$\sigma^2_{U1-U2} = \sigma^2_{U1} + \sigma^2_{U2} . \qquad \mu_{U1-U2} = \mu_{U1} - \mu_{U2} ;$$

$\mu_{U1} = (3 + 7 + 8)/3 = 6$; $\mu_{U2} = (2 + 4)/2 = 3 \Rightarrow$

$\mu_{U1} - \mu_{U2} = 6 - 3 = 3$

$\mu_{U1-U2} = (1 + 5 + 6 - 1 + 3 + 4)/6 = 3$

$\sigma^2_{U1} = (3^2 + 7^2 + 8^2)/3 - 6^2 = 14/3$;

$\sigma^2_{U2} = (2^2 + 4^2)/2 - 3^2 = 1 \Rightarrow \sigma^2_{U1} + \sigma^2_{U2} = 17/3$

$\sigma^2_{U1-U2} = (1^2 + 5^2 + 6^2 + 1^2 + 3^2 + 4^2) / 6 - 3^2 =$

$(1 + 25 + 36 + 1 + 9 + 16) / 6 - 9 = 17/3$

U_1			**$U_1 - U_2$**	
8	7	3		
6	**5**	**1**	2	U_2
4	**3**	**-1**	4	

[1] م سبياجال 1985 ص 180

المبحث 5. توزيع المعاينة للتباين وتوزيع المعاينة لنسبة تبايني عينتين

توزيع المعاينة للتباين
توزيع المعاينة لنسبة تباينين

1 توزيع المعاينة للتباين

(أ) حالة المعاينة بالإرجاع

مسألة: أحسب تباين المجتمع في المسألة 1، أحسب القيمة المتوقعة لتباين العينة المسحوبة بالإرجاع من خلال متوسط تباينات العينات الممكنة، قارن بين تباين المجتمع والقيمة المتوقعة لتباين العينة.

التباينات الممكنة S^2_i				
12,3	6,25	4	1	0
6,25	2,25	1	0	1
2,25	0,25	0	1	4
1	0	0,25	2,25	6,25
0	1	2,25	6,25	12,3

$(\sum_i S^2_i)/25 = 73/25 = 2.92 \quad \Rightarrow E(S^2) = 2.92$

$\sigma^2 = E(X^2) - E(X)^2$

$= [(1 + 9 + 25 + 36 + 64)/5] - 4.6^2 = (135/5) - 21 = 5.84$

$\sigma^2 (1/n) \quad = \quad E(S^2) = 2.92 = 5.84/2$

نظرية 8 : إذا كانت م ع تمثل مجتمع ما وS^2 متغيرة ع تمثل تباين عينة مسحوبة بالإرجاع (أو بدون إرجاع من مجتمع غير محدود) حجمها n، فإن : $E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)$

(عند $n \geq 30$: $E(S^2) \approx \sigma^2$)

البرهان:

$$E(S^2) = E\left[\frac{1}{n}\sum_i (x_i - \bar{x})^2\right] = E\left[\frac{1}{n}\sum_i \ x_i{}^2 - \bar{x}^2\right] = \frac{1}{n}\sum_i \ E(x_i{}^2) - E(\bar{x}^2)$$

$$= \frac{1}{n}\sum_i \left(V(X) + \mu^2\right) - [V(\bar{x}) + E(\bar{x})^2] = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \sigma^2(1 - \frac{1}{n}) = \sigma^2\left(\frac{n-1}{n}\right)$$

ملاحظة: من النظرية نجد أن: $E\left(S^2\frac{n}{n-1}\right) = \sigma^2$ ونقول عن $S^2\frac{n}{n-1}$ أنه مقدر "غير منحرف" ل σ^2 **و**يرمز له ب ‘S^2 حيث:

$$\hat{S}^2 = S^2\frac{n}{n-1}$$

نظرية 9 : إذا أخذنا عينات عشوائية حجمها n من مجتمع طبيعي، فإن : $\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$

مثال : ليكن مجتمع طبيعي حجمه 100 نسحب منه عينة حجمها n = 16 . ما هو احتمال أن يكون تباين العينة S^2 أقل من أو يساوي 10 علما أن تباين المجتمع 80.

$$P(S^2 \le 10) = P(\chi^2_{15} \le 2) = P(S^2 \frac{16}{80} \le X \sim N(\mu,\sigma)) \Rightarrow \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

من الجدول $P(X^2_{15} \le 2) < 0.005$

ب) حالة المعاينة بدون إرجاع

مسألة: في المسألة 1 أحسب تباين المتوسطات الممكنة للعينة σ^2_m في حالة المعاينة بدون إرجاع، قارن بين تباين المجتمع وتباين المتوسطات الممكنة للعينة.

التباينات الممكنة S^2_i			
1			
4	1		
6,25	2,25	0,25	
12,3	6,25	2,25	1

$= \sigma^2 * [(n-1)/ n] [N/ (N-1)]$

$(\sum_i S^2_i) = 36.5$; $(\sum_i S^2_i)/10 = 3.65$ => $E(S^2) = \mathbf{3.65}$

$\sigma^2 = E(X^2) – E(X)^2$

$= [(1 + 9 + 25 + 36 + 64)/5] - 4.6^2 = 5.84$

$E(S^2) = 3.65 = 5.84*(5/4) (1/2)$

نظرية 10 : إذا كانت م ع تمثل مجتمع ما محدود وS^2 متغيرة ع تمثل تباين عينة نفادية مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لتباين العينة تكتب: $E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right)$

(عندما يكون N كبير جدا N/ (N-1) تؤول إلى 1)

2 توزيع المعاينة لنسبة تباينين

رأينا في الفصل السابق أن: $X = \frac{X_1/v_1}{X_2/v_2} \sim F_{v1,\ v2}$ في حالة المتغيرتان العشوائيتان مستقلتان و$X_1 \sim \chi^2_{v_1}$ و $X_2 \sim \chi^2_{v_2}$. من النظرية 9 نستنتج ما يلي:

نظرية 12 : ليكن لدينا مجتمعان طبيعيان تبايناهما σ^2_1 , σ^2_2 . نسحب من المجتمعين عينتين عشوائيتين حجمهما على التوالي n_1 , n_2 :

$$F = \frac{\left[\frac{S_1^2 n_1}{n_1-1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2-1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \to F_{n_1-1;\ n_2-1}$$

مثال[1] **.** عينتين حجمهما 8 و10 مسحوبتين من مجتمعين طبيعيين تبايناهما على التوالي 20 و36. ما احتمال أن يكون تباين الأولى أكبر من ضعف تباين الثانية؟

[1] م سبيجال 1985 ص 186.

$$P(S_1^2 > 2S_2^2) = P(\frac{S_1^2}{S_2^2} > 2) = P\left(\frac{S_1^2\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{S_2^2\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} > 2\frac{\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} \right) = P\left(\frac{S_1^2\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{S_2^2\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} > 2\frac{\left(\frac{8}{7}\right)\frac{1}{20}}{\left(\frac{10}{9}\right)\frac{1}{36}} \right) =$$

$$= P(F_{7,9} > 3.7)$$

من الجدول نجد $0.05 > P(F_{7,9} > 3.7) > 0.01$ و في الحقيقة $P(F_{7,9} > 3.7) = 0.036$

3 ملحق

أ) الانحراف المعياري لتوزيع المعاينة للتباين

نظرية 11 : إذا كانت X م ع تمثل مجتمع ما وS^2 متغيرة ع تمثل تباين عينة مسحوبة من ذات المجتمع، فإن:

$$\sigma_{S^2} = \begin{cases} \sigma^2\sqrt{2/n} & \text{si } X \sim N \\ \sqrt{\dfrac{\mu^4 - \sigma^4}{n}} & \text{sinon} \end{cases}$$

من أجل $n \geq 100$، توزيع S^2 يقترب كثيرا من التوزيع الطبيعي.

ب) الانحراف المعياري لتوزيع المعاينة للانحراف المعياري

$$\sigma_S = \begin{cases} \dfrac{\sigma}{\sqrt{2/n}} & \text{si } X \sim N \text{ ou } X \approx N \\ \sqrt{\dfrac{\mu^4 - \sigma^4}{4n\sigma^2}} & \text{sinon} \end{cases}$$

من أجل $n \geq 100$، توزيع S يقترب كثيرا من التوزيع الطبيعي و$\mu_S \approx S$

4 خلاصة

الجدول التالي يلخص ما ورد في النظريات السابقة من 6 إلى 10 .

إحصائية العينة	المجتمع	المعاينة	الخاصية
النسبة	مجتمع موزع طبيعيا غير محدود		$E(p') = \mu_{p'} = p \quad ; \quad \sigma^2{}_{p'} = \frac{pq}{n}$
		$n \geq 30$	$p' \approx N(p, \sigma_{p'})$
	مجتمع طبيعي محدود	المعاينة نفاديه	لحساب $\sigma_{p'}$ نضرب في معامل الإرجاع.
الفرق بين إحصائيتين ما.	مجتمع ما	سحب بالإرجاع	$\mu_{S - S2} = \mu_{S1} - \mu_{S2}$ $\mu_{S1 + S2} = \mu_{S1} + \mu_{S2}$ $\sigma^2_{S1 - S2} = \sigma^2_{S1} + \sigma^2_{S2}$ $\sigma^2_{S1 + S2} = \sigma^2_{S1} + \sigma^2_{S2}$
		n_2 و $n_1 \geq 30$	$\mu_{m1 - m2} \approx N(0, 1)$
التباين	مجتمع ما وتباين عينة S^2	سحب بالإرجاع (أو بدون إرجاع من مجتمع غير محدود) حجمها n	$E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)$
		$n \geq 30$	$E(S^2) \approx \sigma^2$
	مجتمع طبيعي	حجمها n	$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$
	مجتمع ما محدود وS^2 تمثل تباين العينة	عينة نفادية	$E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right)$
		N كبير جدا	$N/(N-1)$ تؤول إلى 1
نسبة تباينين	مجتمعان طبيعيان تبايناهما $\sigma^2 1$, $\sigma^2 2$	عينتين عشوائيتين حجمهما على التوالي n_1 , n_2	$F = \frac{\left[\frac{S_1^2 n_1}{n_1 - 1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2 - 1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} \to F_{n_1 - 1;\, n_2 - 1}$

الفصل VIII. نظرية التقدير

مفاهيم أساسية طرق التقدير بمجال طرق تأسيس المقدر

في الفصل السابق درسنا من خلال مجموعة من النظريات العلاقة الرياضية بين معالم العينة والمعالم المناظرة لها في المجتمع مثل المتوسط، التباين، النسبة...كما درسنا العلاقة بين شكل توزيع المجتمع وشكل التوزيع الاحتمالي لمعالم العينة. تظهر هذه العلاقات كتوصيف لخصائص العينة ومعالمها ولكنها تستخدم أكثر لتقدير خصائص ومعالم المجتمع محل الدراسة، وهذا ما سنتعرف عليه في هذا الفصل.

المبحث 1. مفاهيم أساسية

بعض خصائص المقدر
التقدير النقطي، التقدير بمجال

1 بعض خصائص المقدر[1]

لتقدير معلمة من معالم مجتمع محل دراسة، نحتاج إلى اختيار الإحصائية المناسبة في العينة لتقدير هذه المعلمة. غالبا ما تكون المعلمة المناظرة في العينة هي أحسن مقدر، كأن نقدر متوسط المجتمع μ من خلال متوسط العينة μ_m. تسمى الإحصائية المستخدمة في التقدير المقدر.

(أ) المقدر غير المتحيز

نقول عن إحصائية ما بأنها مقدر غير متحيز sans biais لمعلمة المجتمع إذا كان متوسطها أو توقعها الرياضي مساويا لمعلمة المجتمع.

مثال: نقول عن متوسط العينة m أنه مقدر غير متحيز لمتوسط المجتمع μ لأن $E(m) = \mu$. في المقابل نسمي الإحصائية S^2 في معاينة بالإرجاع أنها مقدر متحيز ل σ^2 لأن $E(S^2) = \sigma^2 (n-1)/n \neq \sigma^2$ ، بينما تعتبر الاحصائية = S'^2 $S^2n/(n-1)$ مقدرا غير متحيز في معاينة بالإرجاع.

(ب) الكفاءة

تتعلق كفاءة (efficacité) مقدر ما بمقدار التباين لتوزيع المعاينة للإحصائية، فإذا كان لمقدرين (إحصائيتين) نفس المتوسط نقول عن المقدر ذو توزيع المعاينة الأقل تباينا أنه الأكثر كفاءة.

مثال: لكل من توزيعي المعاينة للمتوسط والوسيط نفس المتوسط هو متوسط المجتمع μ ، لكن يعتبر المتوسط m مقدرا أكثر كفاءة لمتوسط المجتمع μ من الوسيط لأن تباين توزيع المعاينة للمتوسطات $V(m) = \sigma^2/n$ أقل من تباين توزيع المعاينة للوسيط :

$$V(méd) = \sigma^2\pi/2n = (\sigma^2/n)\ (3.14159/2) \quad > \quad \sigma^2/n.$$

من البديهي أن استخدام مقدرات فعالة وغير متحيزة هو الأفضل، إلا أنه قد يلجأ لمقدرات أخرى لسهولة الحصول عليها.

[1] سبياجال 1985، ص 204.

(ج) **التقارب convergeance**

نقول عن مقدر أنه متقارب إذا كان يؤول إلى قيمة المعلمة المقدرة عندما يؤول حجم العينة إلى ما لا نهاية.

مثال: يعتبر متوسط العينة مقدرا متقاربا لمتوسط المجتمع لأن:

$$E(m)=\mu \quad , \quad V(m)=\frac{\sigma^2}{n}\xrightarrow[n\longrightarrow\infty]{}0.$$

2 التقدير النقطي والتقدير بمجال.[1]

قد نحتاج إلى تقدير لمعلمة مجتمع بقيمة واحدة ونقول عن هذا التقدير أنه **تقدير نقطي**، و أحيانا أخرى نحتاج إلى تقدير معلمة المجتمع بنقطتين يحددان مجال لقيمة المعلمة ونقول عن هذا النوع من التقدير أنه **تقدير بمجال**.

مثال : إذا قدرنا دخل الأسرة في منطقة ما ب 18000 دج، نكون قد قدرنا دخل الأسرة تقديرا نقطيا. يكون تقديرنا بمجال إذا قلنا مثلا أن الدخل يساوي 18000 ± 2000 أي أنه يتراوح بين 16000 و20000دج.

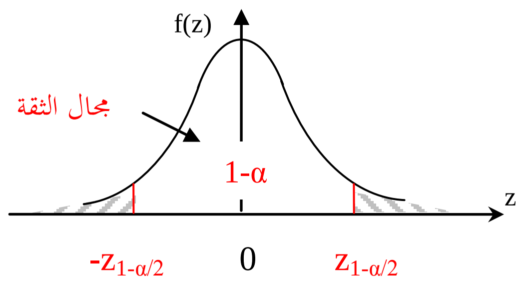
أ) درجة التأكد

لكي يكون التقدير علميا ينبغي تقييم احتمال أن تكون المعلمة تنتمي فعلا إلى المجال المحدد، لذلك نلحق بالمجال ما يسمى بدرجة أو مستوى الثقة، ويرمز له ب p. الاحتمال المعاكس يسمى احتمال الخطأ ويرمز له ب α ، ويسمى أيضا ""مستوى المعنوية".

مثال: دخل الأسرة في المنطقة (أ) ينتمي إلى المجال [16000، 20000] بمستوى معنوية 5 % أي بمستوى ثقة 95 % . وتسمي الحدود 16000 و20000 **حدود الثقة**.

(ب) تعيين حدود مجال الثقة

تحدد حدود الثقة من خلال معاملات الثقة التي بدورها تحدد من خلال مستوى المعنوية (مستوى الثقة). ففي حالة استخدام التوزيع الطبيعي للتقدير تكون القيمتين ± 1.96 معاملات الثقة من أجل مستوى ثقة 95% بينما القيمتين ±2.58 تمثلان معاملات الثقة من أجل مستوى ثقة 99 % .



رسم 21 مجال الثقة للتوزيع الطبيعي

مثال: ليكن μ_s وσ_s متوسط وانحراف معياري توزيع المعاينة لإحصائية ما s حيث $\mu_s=\mu$. إذا كان توزيع المعاينة ل s توزيعا طبيعيا (كما هو الحال بالنسبة لأغلب الإحصائيات عندما $(n \geq 30)$) فإننا نقدر مثلا وبالنظر إلى توزيع s أن: القيمتين $\mu_s \pm 1.96\sigma_s$ تمثلان **حدود الثقة** ب 95 %، و $\mu_s \pm 2.58\sigma_s$ حدود الثقة ب 99 %.

[1] المرجع السابق

في حالة التوزيع الطبيعي يرمز لحدود الثقة ب Z_c أو $Z_{1-\alpha/2}$ (أنظر الرسم).

المبحث 2. التقدير بمجال

كيفية تعيين مجال الثقة للمتوسط
كيفية تعيين مجال الثقة للنسبة
كيفية تعيين مجال الثقة للتباين
كيفية تعيين مجال الثقة لنسبة تباينين

1 مجال الثقة للمتوسط

يقدر متوسط المجتمع μ من خلال الإحصائية m.

(أ) تقدير μ باستخدام التوزيع الطبيعي

نستخدم التوزيع الطبيعي لتحديد مجال الثقة إذا علمنا أن المجتمع الذي سحبت منه العينة يتبع التوزيع الطبيعي.

وفي حالة العينة الممتدة $(n \geq 30)$ يمكن كذلك الاستفادة من نظرية النهاية المركزية[1] أن m تتبع التوزيع الطبيعي.

تكتب حدود مجال الثقة كما يلي:

$$m \pm z_c \cdot \frac{S}{\sqrt{n-1}} \quad \text{أو} \quad m \pm z_c \frac{S'}{\sqrt{n}}$$

وفي حالة σ مجهول:

$$m \pm z_c \cdot \frac{\sigma}{\sqrt{n}}$$

و نستخدم هذه الصيغة إلا إذا كان المجتمع محدود (ذا حجم N) والمعاينة نفادية حيث تصبح الصيغة كالآتي:

$$m \pm z_c \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

إلا أنه غالبا ما يكون الانحراف المعياري للمجتمع σ مجهولا، ولذلك نعوض σ في الصيغ السابقة بالمقدر S' أو S.

الجدول الآتي يبين قيم z_c التي تمثل حدود مجال الثقة بحسب مستوى الثقة :

مستوى الثقة $1-\alpha$	0.99	0.98	0.95	0.90	0.8	0.5
مستوى المعنوية α	0.01	0.02	0.05	0.10	0.2	0.5
$1-\alpha/2$	0.995	0.99	0.975	0.95	0.9	0.75
$Z_{1-\alpha/2}$	82.5	2.326	1.96	1.645	1.282	0.674

مثال : نقدر أن μ يوجد داخل المجال $m \pm 1.96\sigma_m$ بمستوى ثقة 95% (0.95) أي بمستوى معنوية 5 % (0.05)، وداخل المجال $m \pm 2.58\sigma_m$ بمستوى ثقة 99% أي بمستوى معنوية 0.01...

(ب) تقدير μ باستخدام التوزيع t :

في حالة العينة الصغيرة $(n < 30)$ و σ مجهول نستخدم توزيع ستيودنت لتحديد مجالات الثقة ل μ. مثلا القيم- $t_{0.975}$ ؛ $t_{0.975}$ تحد 95 % من المساحة تحت المنحنى ونقول أن $t_{0.975}$; $-t_{0.975}$ تمثل القيم الحرجة أو معاملات الثقة عند مستوى ثقة 95 % ونكتب:

[1] التي تخص في الحقيقة توزيع مجموع قيم العينة –في حالة كون العينة كبيرة بما فيه الكفاية– وليس المتوسط.

$$-t_{0.975} < \frac{m-\mu}{\hat{S}/\sqrt{n}} < t_{0.975}$$

ومنه نستخلص مجال الثقة ل μ كما يلي:

$$m - \ t_{0.975}\cdot\frac{\hat{S}}{\sqrt{n}} < \ \mu \ < \ m \ + \ t_{0.975}\cdot\frac{\hat{S}}{\sqrt{n}}$$

2 مجال الثقة للنسبة

(أ) حالة المجتمع غير محدود أو المعاينة غير نفادية و العينة الممتدة ($n \geq 30$) :

لتكن s إحصائية تمثل نسبة "نجاحات" في عينة ذات حجم $n \geq 30$ مستخرجة من مجتمع ثنائي حيث p هي نسبة النجاحات. تستعمل التوزيع الطبيعي لتقدير p فنعين حدود الثقة ل p كما يلي: $p' \pm z_c\sigma_p$ أين p' نسبة النجاحات في العينة،

نعلم من الفصل السابق أن $\sigma_p = \sqrt{\frac{pq}{n}}$ ومنه يحدد مجال الثقة ل p كما يلي:

$$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}}$$

(ب) في حالة كون المجتمع محدود ذا حجم N والمعاينة نفادية:

$$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

3 مجال الثقة للتباين

لتقدير التباين والانحراف المعياري لمجتمع بمجال ثقة نستعمل الخاصية : $\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$.

مثال: مجال الثقة ب95 % يحدد كما يلي:

$$\chi^2_{0.025} \leq \frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \leq \chi^2_{0.975}$$

ومنه نستنتج مجال الثقة ل σ كما يلي:

$$\frac{\sqrt{n}S}{\chi_{0.975}} \leq \ \sigma \ \leq \frac{\sqrt{n}S}{\chi_{0.025}} \qquad ou \qquad \frac{\sqrt{n-1}\hat{S}}{\chi_{0.975}} \leq \ \sigma \ \leq \frac{\sqrt{n-1}\hat{S}}{\chi_{0.025}}$$

نظرا لأن توزيع ك2 غير متماثل فإن المجال أعلاه ليس الأمثل، إذ توجد طريقة لتضييق مجال الثقة أكثر إذا لم نشأ أن تكون أطراف المنحنى متساوية، وهذا بخلاف التوزيعات المتماثلة كالطبيعي وستيودنت.

4 مجالات الثقة لنسبة تباينين

رأينا سابقا (نظرية 11 من الفصل 5) أنه إذا كان لدينا مجتمعان طبيعيان تبايناهما σ^2_1 , σ^2_2 وسحبنا منهما عينتين عشوائيتين حجمهما على التوالي n_1 , n_2 فإن :

$$F=\frac{\left[\frac{S_1^2 n_1}{n_1-1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2-1}\right]\frac{1}{\sigma_2^2}}=\frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2}\rightarrow F_{n_1-1;\,n_2-1}$$

إذا يمكن تكوين تقدير بمجال ل F عند مستوى ثقة 0.98 كما يلي:

$$F_{0.01}\leq\frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2}\leq F_{0.99}$$

و من ثم يمكن تقدير النسبة بين تبايني المجتمعين كما يلي:

$$\frac{1}{F_{0.99}}\frac{\hat{S}_1^2}{\hat{S}_2^2}\leq\frac{\sigma_1^2}{\sigma_2^2}\leq\frac{1}{F_{0.01}}\frac{\hat{S}_1^2}{\hat{S}_2^2}$$

5 خلاصة

لتقدير إحصائية مجتمع نستخدم نظريات توزيع المعاينة. هذه النظريات تتناول خصائص إحصائيات العينة من متوسط العينة، النسبة في العينة، ... و علاقتها بالإحصائيات المناظرة لها في المجتمع.

جدول1 توزيع المعاينة للمتوسطات حسب طبيعة توزيع المجتمع، معلومية التباين و حجم العينة.

قانون المجتمع	**تباين المجتمع (σ^2)**	**n**	$\sigma_{\bar{x}}$	**قانون** $\bar{x}$
طبيعي	معلوم	n < 30 أو n ≥ 30	σ/√n	N(μ ; σ/√n)
	غير معلوم	n ≥ 30	S’/√n	N(μ ; S’/√n)
		n < 30	S’/√n	$t_{\alpha;\, n-1}$
غير معلوم	معلوم	n ≥ 30	σ/√n	N(μ ; σ/√n)
	غير معلوم	n ≥ 100	S’/√n	N(μ ; S’/√n)

جدول 2 تحديد مجال الثقة للنسبة، للتباين وللنسبة بين تباينين

المجتمع	**التوزيع الاحتمالي للإحصائية**	**مجال الثقة**
مجتمع غير محدود أو معاينة غير نفادية و عينة ممتدة (n ≥ 30)	التوزيع الطبيعي	$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}}$
مجتمع محدود ذا حجم **N** والمعاينة نفادية	التوزيع الطبيعي	$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
غير معلوم	$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$	$\frac{\sqrt{n}S}{\chi_{0.975}} \leq \sigma \leq \frac{\sqrt{n}S}{\chi_{0.025}}$ أو $\frac{\sqrt{n-1}\hat{S}}{\chi_{0.975}} \leq \sigma \leq \frac{\sqrt{n-1}\hat{S}}{\chi_{0.025}}$
مجتمعين طبيعيين، أو عينتين مسحوبتين من مجتمع طبيعي واحد.	$F = \frac{\left[\frac{S_1^2 n_1}{n_1-1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2-1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \rightarrow F_{n_1-1;\, n_2-1}$	مثلا عند مستوى ثقة 0.98: $\chi^2_{0.025} \leq \frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \leq \chi^2_{0.975}$ $\frac{1}{F_{0.99}}\frac{\hat{S}_1^2}{\hat{S}_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1}{F_{0.01}}\frac{\hat{S}_1^2}{\hat{S}_2^2}$

6 ملحق. مجالات الثقة للفروق والمجاميع

إذا كانت s_1 وs_2 إحصائيتا معاينة لها توزيع يقترب من التوزيع الطبيعي، والعينتان مستقلتان، تكتب حدود الثقة للفروق بين المعالم التي تمثلها الإحصائيتين كما يلي:

$$S_1 - S_2 \ \pm\ z_c \cdot \sigma_{S_1 - S_2} = S_1 - S_2 \ \pm\ z_c \cdot \sqrt{\sigma^2{}_{S_1} + \sigma^2{}_{S_2}}$$

في حالة المجموع :

$$S_1 + S_2 \ \pm\ z_c \cdot \sigma_{S_1 - S_2} = S_1 + S_2 \ \pm\ z_c \cdot \sqrt{\sigma^2{}_{S_1} + \sigma^2{}_{S_2}}$$

مثال: إذا كانت الإحصائيتان هما متوسطا عينتين مستقلتين، مسحوبتين من مجتمعين غير محدودين، نحدد مجال الثقة للفرق (و للمجموع) بين متوسطي المجتمعين $\mu_1 - \mu_2$ كما يلي :

$$m_1 - m_2 \ \pm\ z_c \cdot \sigma_{m_1 - m_2} = m_1 - m_2 \ \pm\ z_c \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

مثال 2: إذا كانت الإحصائيتان هما نسبتان في عينتين مستقلتين، مسحوبتان من مجتمعين غير محدودين :

$$p'_1 - p'_2 \ \pm\ z_c \cdot \sigma_{p'_1 - p'_2} = p'_1 - p'_2 \ \pm\ z_c \cdot \sqrt{\frac{pq_1}{n_1} + \frac{pq_2}{n_2}}$$

المبحث 3. طرق تأسيس المقدر[1]

طريقة العزوم
طريقة المعقولية العظمى (الاحتمال الأكبر)

أحد الطرق لاختيار مقدر معلمة ما للمجتمع أن نأخذ مباشرة نظيرتها في العينة، وإذا كان هذا المقدر لا يتصف بالخصائص المطلوبة نجري عليه تعديلا (استخدام S'^2 بدلا من S^2 لتقدير σ^2). توجد طرق أخرى لتحديد المقدر الأنسب منها طريقة المعقولية العظمى والتي تدعى أيضا طريقة الاحتمال الأكبر والتي تنسب إلى العالم فيشر وكذا طريقة العزوم.

1 طريقة العزوم

ليكن المطلوب تقدير عدد K من معالم المجتمع : $\theta_1, \theta_2, \ldots, \theta_k$. نكون جملة معادلات عددها K. تتضمن كل معادلة مساواة العزم المرتبط بالأصل من الدرجة k لمتغيرة المجتمع X : $\mu'_k = E(X^k)$، بنظيره لمتغيرة المعاينة x :

$$m'_k = (1/n)\sum_i x_i^k \qquad k = 1, 2, \quad , K$$

مثال: ليكن $X \sim B(20; p)$. تقدير p بطريقة العزوم انطلاقا من عينة يتم كما يلي:

لدينا عدد المعالم المراد تقديرها K = 1 إذا نحتاج إلى معادلة واحدة : $\mu = 20p$.ومنه $p = 20/\mu$، نأخذ إذا كمقدر لp القيمة: p' ونحسبها كما يلي : $p' = m/20$.

في حالة تقدير معلمتين للمجتمع نحتاج أن نستعمل جملة المعادلتين: $\mu'_2 = m'_2 \ ,\ \mu = m$

[1] دراوزبيك 1997 ص.308 .

مثال2 : لتكن $X \sim N(\mu; \sigma^2)$. نسحب عينة ذات متوسط m، وتباين S^2. لتقدير µ وσ^2 نحتاج إلى حل جملة المعادلتين:

$$\begin{cases}\mu'_2 = m'_2\\ \mu = m\end{cases} \text{ or } \begin{cases}\mu'_2 = \mu^2+\sigma^2\\ m'_2 = m^2+S^2\end{cases} \quad \text{la solution est} \quad \begin{cases}\hat{\mu} = m\\ \hat{\sigma}^2 = S^2\end{cases}$$

هذه الطريقة قد تعطي مقدرات متحيزة كما في هذه الحالة.

2 طريقة المعقولية العظمى (طريقة الاحتمال الأكبر)

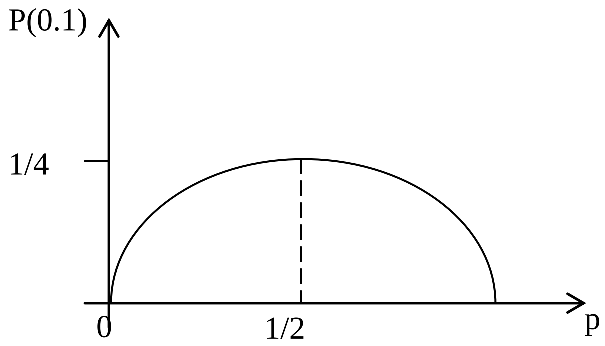
حالة كون متغيرة المجتمع متقطعة : نريد تقدير معلمة θ واحدة للمجتمع، ولدينا عينة غير نفادية (المتغيرا ت التي تمثل قيم المحصل عليها في العينة مستقلة) لها نفس التوزيع للمجتمع. من البديهي أن احتمال تحقق عينة بذاتها مرتبط ب قيمة المعلمة المجهولة : $P(x_1, x_2, \ldots,x_n) = L(\theta)$. هناك قيمة ل θ تعظم احتمال الحصول على العينة المحصل عليها، ونفترض أن تلك القيمة هي الصحيحة بما أن العينة حصلت بالفعل. تتمثل طريقة المعقولية العظمى في البحث عن هذه القيمة. أي البحث عن θ التي تعظم $L(\theta)$ ، حيث :

$$L(\theta) = f(x_1, \ldots, x_n ; \theta) = f(x_1) . f(x_2) \ldots f(x_n).$$

تعتمد طريقة المعقولية العظمى على تعظيم دالة الاحتمال المشتركة $L(\theta)$.

مثال: ليكن $X \sim B(p)$، حيث النجاح هو وجود الخاصية " أ " لدى فرد مسحوب عشوائيا من المجتمع. نرد تقدير p من خلال عينة حجمها 2. ما هي القيمة p' ل p التي تجعل النتيجة 1، 0 هي الأكثر احتمالا؟ أي ما هي p' التي تجعل $p(0.1) = pq$ أكبر ما يمكن؟

من الواضح أن أكبر قيمة ل$p(0.1)$ هي ¼ والقيمة التي تحققها هي $p' = 1/2$ ، وبهذا نجيب على التساؤل.



رسم 22 أقصى قيمة ل P(0,1)

الفصل .IX مفاهيم اختبارات الفروض وتطبيقاتها

اختبار المتوسط
اختبار النسبة واختبار التباين
اختبارات المقارنة بين مجتمعين
اختبار التجانس و اختبار التعديل

في الفصل السابق تناولنا كيفية تقدير معالم المجتمع من خلال بيانات العينة وبعض خصائص المقدر الجيد. في هذا الفصل[1] سنتناول كيفية اختبار فرضيات موضوعة حول معالم مجتمع أو أكثر. يحتاج الدارس أحيانا في مرحلة ما من بحثه إلى اختبار فرضية أو أكثر بخصوص المجتمع المدروس. من أمثلة ذلك: اختبار فرضية بخصوص معدل الدخل في منطقة معينة، اختبار فرضية نسبة شفاء لدواء معين، ... يتم ذلك بصياغة فرضية عن المجتمع المدروس (أو المجتمعات المدروسة) ومن ثم محاولة الحصول على دليل إحصائي ينفي أو يثبت هذه الفرضية وذلك من خلال بيانات عينة (أو أكثر) عشوائية بسيطة. تخص الفرضية أحد معالم المجتمع كالمتوسط، النسبة أو التباين، ونعتمد في إثباتها أو رفضها على خصائص إحصائية معاينة مختارة. من أجل ذلك يعتمد هذا الدرس، كما هو الحال بالنسبة لدرس التقدير، على درس المعاينة.

المبحث 1. اختبار المتوسط

الاختبار ثنائي الاتجاه للمتوسط
الاختبار أحادي الاتجاه للمتوسط
استخدام S كمقدر لتباين المجتمع في اختبار المتزسط
اختبار المتزسط باستخدام توزيع t

يتناول هذا الاختبار متوسط المجتمع (μ)، مثل متوسط الدخل، متوسط وزن منتج معين، .. ويؤكد اختبار المتوسط فرضية مساواته لقيمة ما μ_0. و للقيام بالاختبار نستخرج عينة عشوائية نحسب فيها المتوسط m ثم نستخدم التوزيع الاحتمالي ل m لقياس قرب أو بعد هذه القيمة من μ_0.

1 اختبار ثنائي الاتجاه للمتوسط.

لنتناول هذا المثال: نريد اختبار فرضية حول متوسط دخل الطالب في السنة الأولى من تخرجه، ولتكن القيمة الافتراضية هي 15000دج كمتوسط للدخل الشهري. نحتاج إلى الخطوات التالية: تحديد الفرضيات، تحديد قاعدة القرار، حساب القيمة الجدولية للمتغيرة، حساب القيمة الفعلية للمتغيرة، اتخاذ القرار.

(أ) **تحديد الفرضيات (الصفرية والبديلة):**

$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0$$

[1] تضمن البرنامج فصلين حول موضوع اختبار الفروض، الأول مفاهيم أساسية والثاني تطبيقات اختبار الفروض؛ غير أننا نرى أن الفصل بين هذين الجزئين سوف يؤدي إلى تكرار التطرق للمفاهيم الأساسية. من جهة أخرى، يصعب شرح المفاهيم الأساسية بمعزل عن التطبيقات أي بمعزل عن بنود المبحث الثاني. لذلك فسوف نخوض مباشرة في المبحث الثاني، أي جزء التطبيقات، وفي أثنائه سنتطرق إلى المفاهيم المذكورة في المبحث الأول.

تسمى الفرضية H_0 **الفرضية الصفرية** أو فرضية العدم، ويؤدي الاختبار إما إلى رفضها ونكتب RH_0 وفي هذه الحالة نقبل **الفرضية البديلة** أو المعاكسة أو عدم رفضها ونكتب $R'H_0$. μ_0 هي القيمة الافتراضية ل μ وهي في هذه الحالة 15000 لذلك نكتب الفرضيات كما يلي:

$$H_0 : \mu = 15000 \leftrightarrow H_1 : \mu \neq 15000$$

عادة ما تكون μ_0 محددة بناءا على بيانات عينة عشوائية بسيطة ($\mu_0 = m$)، وفي هذه الحالة يمكن استخدام الخاصية $m \sim N(\mu, \sigma^2/n)$ لاجراء الاختبار، حيث أنه تحت H_0 فإن : $m \sim N(\mu_0 , \sigma^2/n)$

مما يعني معلومية احتمال أن يكون m قريب إلى درجة ما من μ_0 فمثلا :

$$P(\mu_0 - 1.64(\sigma_m) \leq m \leq \mu_0 + 1.64(\sigma_m)) = 0.90$$
$$P(\mu_0 - 1.96(\sigma_m) \leq m \leq \mu_0 + 1.96(\sigma_m)) = 0.95$$
$$P(\mu_0 - 2.58(\sigma_m) \leq m \leq \mu_0 + 2.58(\sigma_m)) = 0.99$$

وبصفة عامة نكتب:

$$P[\mu_0 - z_{1-\alpha/2}(\sigma_m) \leq m \leq \mu_0 + z_{1-\alpha/2} (\sigma_m)] = 1-\alpha$$

أو حسب الكتابة الأكثر شيوعا:

$$P(-z_{1-\alpha/2} \leq \frac{m-\mu_0}{\sigma_m} \leq z_{1-\alpha/2}) = 1-\alpha$$

حيث:

- $(m - \mu_0)/\sigma_m$: (**متغيرة القرار**) هي المتغيرة المعيارية ل m ونرمز لها ب z_c، حيث $z \sim N(0, 1)$.
- σ_m تحدد كما يلي: $\sigma_m = \sigma/\sqrt{n}$ في حالة المعاينة بالإرجاع (أو $n \leq 0.05N$) و $\sigma_m = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ في الحالة المعاكسة.
- $1 - \alpha/2$: المساحة على يسار z .
- n : حجم العينة.

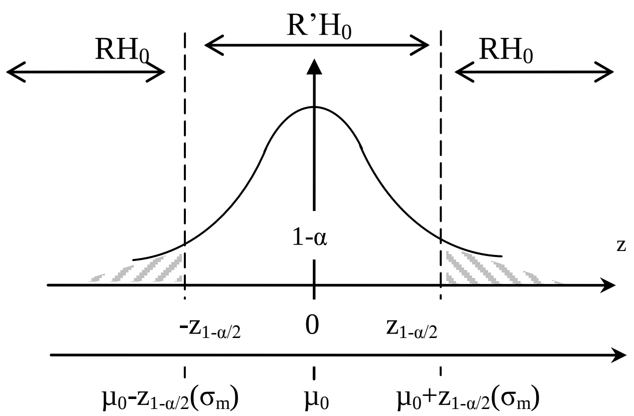
يمكن إذا كان m خارج المجال $1-\alpha$، أن نرفض الفرضية الصفرية التي حدد على أساسها هذا المجال ونقبل بالتالي الفرضية البديلة.

تسمى هذه (الخطة) قاعدة القرار.

(ب) تحديد قاعدة القرار

تكتب قاعدة القرار في المثال الذي بين أيدينا، وهي قاعدة **اختبار ثنائي الاتجاه** (أنظر الشكل 1) ، كما يلي:

$$\begin{cases} RH_0 \ si \left| \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \right| > z_{1-\alpha/2} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases} \quad \text{أو} \quad \begin{cases} RH_0 \ si \ z_c = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \notin \left[-z_{1-\alpha/2}; \ z_{1-\alpha/2} \right] . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$



رسم 23 منطقتي القبول و الرفض في حالة قاعد القرار الثنائية

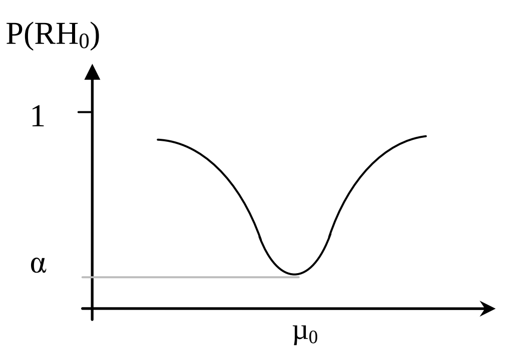
تتضمن هذه الخطة مخاطرة تتمثل في الوصول إلى قرار خاطىء: فقد تكون الفرضية H_0 صحيحة بينما تقودنا قيمة m المحصلة إلى رفضها، ويسمى هذا **الخطأ من النوع الأول**، واحتماله α ، ويكتب : $P(RH_0 / H_0) = \alpha$،

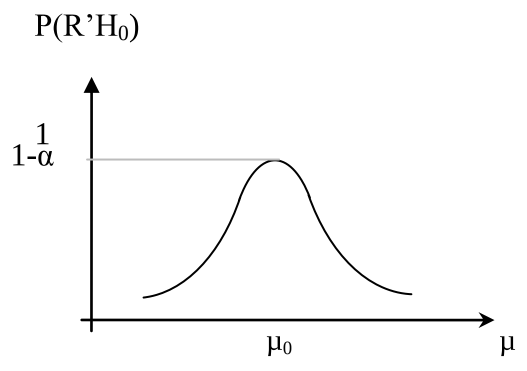
و قد تقودنا قيمة m إلى قبول H_0 فيما هي خاطئة، ويسمى هذا **الخطأ من النوع الثاني** واحتماله α -1 ويكتب :

$P(R'H_0 / H_1) = 1\text{-}\alpha$

و يمكن تقليص احتمال أحد الخطأين على حساب الثاني، ولكن لا يمكن تقليص احتمال كلا الخطأين معا إلا بزيادة حجم العينة.

و يقيس احتمال رفض الفرضية الصفرية $P(RH_0)$ قوة الاختبار (أنظر الشكل 2) فيما يقيس احتمال قبولها $P(R'H_0)$ فعالية الاختبار (أنظر الشكل 2). ويتوقف كلا الاحتمالين على القيمة الحقيقية ل μ .





منحنى القوة (2)

رسم25 (1) منحنى الفعالية

(ج) حساب z الجدولية:

ويرمز لها ب z_t حيث، وهي المشار إليها في قاعدة القرار (الشكل الثاني)، وفي حالتنا (اختبار ثنائي بمستوى معنوية 5%) :

$$z_t = z_{1-\alpha/2} = z_{1-0.05/2} = z_{1-0.025} = z_{0.975}$$

ومن الجدول نجد أن $z_{0.975} = 1.96$.

(د) حساب z الفعلية:

ويرمز لها ب z_c ,وهي المتغير المعيارية ل m (أنظر قاعدة القرار الشكل الأول) :

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{15800 - 1500}{1500/\sqrt{100}} = 5.33$$

(هـ) القرار:

نقرر قبول أو رفض H_0 حسب قاعدة القرار. وفي حالتنا نرفض H_0 لأن $z_c > z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج.

2 الاختبار أحادي الاتجاه للمتوسط.

يتميز الاختبار الثنائي عن الأحادي في الفرضية البديلة التي هي عدم مساواة في الاختبار الثنائي وأكبر تماما أو أصغر تماما (حسب الحالة) في الاختبار الأحادي، وهذا يترتب عليه تغيير في قاعدة القرار.

(أ) الاختبار أحادي الاتجاه من اليمين.

لنرجع إلى المثال السابق مع تغيير محدد هو أننا نريد اختبار ما إذا كان متوسط الدخل للخريج 15000دج أم أكثر (اختبار من اليمين).

أ- **الفرضيات** : $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu > \mu_0$

في هذه الحالة $\mu_0 = 15000$ لذلك نكتب : $H_0 : \mu = 15000 \leftrightarrow H_1 : \mu > 15000$

ب- **قاعدة القرار:**

$$\begin{cases} RH_0 \ si \ \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} > z_{1-\alpha} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

ج- **حساب z الجدولية**: (اختبار على اليمين بمستوى معنوية 5%) : $z_t = z_{1-\alpha} = z_{1-0.05} = z_{0.95}$

ومن الجدول نجد أن $z_{0.95} = 1.645$

د- **حساب z الفعلية:**

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{15800 - 1500}{1500/\sqrt{100}} = 5.33$$

هـ- **القرار**: نرفض H_0 لأن $z_c > z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج وإنما هو أكبر.

(ب) الاختبار أحادي الاتجاه من اليسار

نعود إلى مثالنا ونفترض أن متوسط العينة كان 14200دج ونريد أن نختبر ما إذا كان متوسط الدخل مساوي أم أقل من 15000دج.

أ- **الفرضيات** : $H_0 : \mu = 15000 \leftrightarrow H_1 : \mu < 15000$

ب- **قاعدة القرار**:

$$\begin{cases} RH_0 \ si \ \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} < -z_{1-\alpha} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

ج- **حساب z الجدولية**: (اختبار على اليسار بمستوى معنوية 5 %) :

$$z_t = -z_{1-\alpha} = -z_{1-0.05} = -z_{0.95} = -1.645$$

د- **حساب z الفعلية**:

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{14200 - 1500}{1500 / \sqrt{100}} = -5.33$$

هـ- **القرار**: نرفض H_0 لأن $z_c < z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف أقل من 15000دج .

3 استخدام S كمقدر ل σ في اختبار المتوسط.

في الأمثلة السابقة افترضنا أن σ معلوم، في الواقع غالبا ما يكون الانحراف المعياري مجهولا ونحتاج بالتالي إلى استخدام الانحراف المعياري للعينة (S) عند حساب σ_m (أنظر درس التقدير)، حيث نعوض العبارة

$\sigma m = \sigma/\sqrt{n}$ ب $\sigma_m = \dfrac{S}{\sqrt{n-1}}$ أو $\sigma_m = \dfrac{S'}{\sqrt{n}}$

مثال: في المثال السابق نفترض أن الانحراف المعياري للدخل الشهري للطالب مجهول، لكن الانحراف المعياري للعينة S = 1600. كيف يمكن اختبار ما إذا كان الدخل الشهري أقل من 15000دج؟

الخطوات أ، ب ، ج تبقى بدون تغيير.

د- **حساب z الفعلية**:

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_0}{S / \sqrt{n-1}} = \frac{14200 - 1500}{1600 / \sqrt{99}} = -4.97$$

هـ- **القرار**: نرفض H_0 لأن $z_c < z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج وإنما هو أقل.

4 استخدام التوزيع t في اختبار المتوسط.

في حالة $n < 30$ و σ (الانحراف المعياري للمجتمع) مجهولا، لا يمكن استخدام التوزيع الطبيعي، ولكن لدينا :

$$\frac{\overline{X}-\mu}{S/\sqrt{n-1}} \sim t_{n-1}$$ و تحت H_0 : (μ = μ0) : $$\frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} \sim t_{n-1}$$

يمكن إذا استخدام التوزيع ستيودنت (بشرط أن يكون توزيع المجتمع طبيعيا أو على الأقل جرسي الشكل) .

و تتغير قاعدة القرار تبعا لهذا التغيير فتكتب في حالة الاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ si \left|\frac{\overline{X}-\mu_0}{S/\sqrt{n-1}}\right| > t_{n-1;\,1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليمين:

$$\begin{cases} RH_0 \ si \frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} > t_{n-1,\,1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:

$$\begin{cases} RH_0 \ si \frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} < -t_{n-1;\,1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

5 خلاصة

يتم اختبار الفرضيات من خلال 5 خطوات متتالية وهي:

- تحديد الفرضيات (الصفرية والبديلة)
- تحديد قاعدة القرار
- حساب القيمة الجدولية للمتغيرة
- حساب القيمة الفعلية للمتغيرة
- اتخاذ القرار.

تتحدد كيفية إتمام كل خطوة حسب طبيعة الاختبار (ثنائي أو أحادي الاتجاه)، حسب طبيعة المجتمع و طبيعة و حجم العينة، ... و تسخدم في ذلك نظريات توزيع المعاينة.

المبحث 2. اختبار النسبة واختبار التباين

اختبار النسبة
اختبار التباين

1 اختبار النسبة

يتعلق هذا الاختبار بنسبة مفردات المجتمع التي تتصف بخاصية ما (p)، حيث يؤكد الاختبار أو ينفي صحة فرضية معينة بخصوص قيمة p . يرمز للقيمة الافتراضية ب p_0 وتكتب الفرضية كما يلي: $H_0 : p = p_0$

للقيام بالاختبار نستخدم خصائص p' النسبة في العينة (أنظر توزيع المعاينة للنسبة : نظرية 6).

$$E(p') = \mu_{p'} = p \qquad ; \qquad \sigma^2_{p'} = \frac{pq}{n}$$

عند $n \geq 30$: $p' \approx N(p, \sigma_{p'})$ (نظرية موافر – لابلاس)

استنادا إلى هذه الخصائص وتحت H0 : $p' \approx N\left(p_0 ; \sqrt{\frac{p_0 q_0}{n}}\right)$

و من ثم يمكن تحديد قاعدة القرار بحسب طبيعة الاختباركما يلي:

في حالة الاختبار الثنائي:
$$\begin{cases} RH_0 \ si \left| \dfrac{p' - p_0}{\sqrt{p_0 q_0 / n}} \right| > z_{1-\alpha/2} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليمين:
$$\begin{cases} RH_0 \ si \dfrac{p' - p_0}{\sqrt{p_0 q_0 / n}} > z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:
$$\begin{cases} RH_0 \ si \dfrac{p' - p_0}{\sqrt{p_0 q_0 / n}} < -z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

مثال: تقدر الدوائر الرسمية نسبة المتخرجين الجامعيين الذين يحصلون على عمل في السنة الأولى التي تلي تخرجهم ب 70 % . وجدت دراسة أجريت على عينة من 900 طالب أن نسبة الحصول على عمل 67 % . كيف يمكن اختبار ما إذاكانت النسبة الرسمية صحيحة أم مبالغ فيها، بمستوى معنوية 5 %.

$$H_0 : p = 0.70 \leftrightarrow H_1 : p < 0.70$$

$$\frac{p' - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.67 - 0.7}{\sqrt{0.7(0.3)/900}} \cong -196.34 < -z_{1-0.05} = -1.64$$

ومنه نرفض الفرضية H_0 .

2 اختبار التباين

لاختبار صدقية فرضية بخصوص قيمة تباين مجتمع ما،

$$H_0 : \sigma^2 = \sigma_0^2 \quad \leftrightarrow H_1 : \sigma^2 = \sigma_0^2$$

نستعمل المقدر غير المنحاز $\hat{S}^2 = \dfrac{\sum_i (X_i - \overline{X})^2}{n-1}$. حيث في حالة العينة الكبيرة ($n \geq 50$ في أحسن الأحوال) ،

وتحت H_0 فإن $\mu_4 = E(X - \mu)^4$

$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \approx N(0.1).$$

حيث μ_4 هو العزم المركزي من الدرجة الرابعة. وبهذا الشكل تكتب قاعدة القرار للاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ \ si \left| \dfrac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \right| > z_{1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

وفي حالة μ_4 مجهول يمكن استخدام كمقدر : $m_4 = E(xi - m)^4$.

وإذا كان المجتمع طبيعيا، حيث $\mu_4 = 3\sigma^4$ ، فإن **متغيرة القرار** يمكن أن تكتب كما يلي:

$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sigma_0^2 \sqrt{2/n}} \approx N(0.1).$$

المبحث 3. اختبار المقارنة بين مجتمعين

اختبار تساوي متوسطي مجتمعين
اختبار تساوي تبايني مجتمعين

يتناول هذا الاختبار مقارنة بين مجتمعين من خلال المتوسط أو التباين لكل منهما ... وسنركز هنا على متغيرة القرار، إذ من السهل على الطالب استنتاج كيفية إتمام الخطوات الأخرى على ضوء ما سبق.

1 اختبار تساوي متوسطي مجتمعين

الغرض من الاختبار هو تأكيد أو نفي تساوي متوسطي مجتمعين من خلال عينتين عشوائيتين بسيطتين مستقلتين. تكتب الفرضيات (في حالة الاختبار الثنائي) كما يلي: $H_0 : \mu_1 = \mu_2 \leftrightarrow H_1 : \mu_1 \neq \mu_2$

لتحديد متغيرة القرار نعتمد في الاختبار على متغيرة القرار T أو T' بحسب الحالة (نترك للطالب استنتاج قاعدة القرار)، حيث نميز بين حالة كون تباينا المجتمعين معلومين وحالة كون تباينا المجتمعين مجهولين.

(أ) تباينا المجتمعين معلومين

1- المجتمعين طبيعيين:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0.1)$$

2- مجتمعين ما $(n_1 , n_2 \geq 30)$:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0,1)$$

(ب) تباينا المجتمعين مجهولين

1- المجتمعان طبيعيان:

إذا كان تباينا المجتمعين متساويين

$$T' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

2- مجتمعين ما $(n_1 , n_2 \geq 30)$:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}} \approx N(0,1)$$

مثال: نسحب من مجتمعين طبيعيين متساويي التباين عينتين حجم الأولى 18 وحجم الثانية 21. وجدنا النتائج التالية:

$m_1 = 81, m_2 = 76, S^2_1 = 9, S^2_2 = 8.$ كيف يمكن اجراء اختبار تساوي متوسطي المجتمعين بمستوى معنوية 5 %.

$H_0 : \mu_1 = \mu_2 \leftrightarrow H_1 : \mu_1 \neq \mu_2$

$$T' = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{81-76}{\sqrt{\frac{18(9)+21(8)}{18+21-2}\left(\frac{1}{18} + \frac{1}{21}\right)}} \cong 5.43$$

$$t_{0.975;37} \cong 2.336 < 5.43 \qquad \Rightarrow RH_0$$

2 اختبار تساوي تبايني مجتمعين

الغرض من الاختبار هو تأكيد أو نفي تساوي تباينا مجتمعين من خلال عينتين عشوائيتين بسيطتين مستقلتين.

تكتب الفرضيات (في حالة الاختبار الثنائي)كما يلي: $H_0 : \sigma^2_1 = \sigma^2_2 \leftrightarrow H_1 : \sigma^2_1 \neq \sigma^2_2$

نعتمد في الاختبار على متغيرة القرار T أو 'T بحسب الحالة، حيث نميز بين حالة كون المجتمعين طبيعيين أم غير ذلك.

(أ) مجتمعين طبيعيين

1- الحالة العامة:

$$T' = \frac{\hat{S}_1^2}{\hat{S}_2^2} \sim F_{n_1-1;\, n_2-1}$$

2- في حالة $n_1 , n_2 \geq 30$

$$T = \frac{1}{2}\ln\left(\frac{\hat{S}_1^2}{\hat{S}_2^2}\right) / \sqrt{\frac{1}{2}\left(\frac{1}{n_1-1} + \frac{1}{n_2-1}\right)} \approx N(0;1)$$

(ب) مجتمعين ما $(n_1, n_2 \geq 50)$

1- $\mu_4^{(1)} ; \mu_4^{(2)}$ معروفين :

$$T = (\hat{S}_1^2 - \hat{S}_2^2) \,/\, \sqrt{\frac{\mu_4^{(1)} - \hat{S}_1^4}{n_1} - \frac{\mu_4^{(2)} - \hat{S}_2^4}{n_2}} \approx N(0;1)$$

2- في حالة $\mu_4^{(1)} ; \mu_4^{(2)}$ غير معروفين : نعوض μ_4 ب m_4 .

مثال : نسحب من مجتمعين طبيعيين عينتين حجم الأولى 18 وحجم الثانية 21. وجدنا النتائج التالية:

$m_1 = 81, m_2 = 76, S^2_1 = 9, S^2_2 = 8.$ كيف يمكن اجراء اختبار تساوي متوسطي المجتمعين بمستوى معنوية 5 % ؟

$H_0 : \sigma^2_1 = \sigma^2_2 \leftrightarrow H_1 : \sigma^2_1 \neq \sigma^2_2$

$S'^2_1 = S^2_1 * n_1 / (n_1-1) = 9\ (18)/17 \approx 9.53$; $S'^2_2 = S^2_2 * n_2 / (n_2-1) = 8\ (21)/20 = 8.4$

$S'^2_1 / S'^2_2 \approx 1.135$; $F_{0.05\ ;17\ ;20} \approx 2.17$

$T < F_{\alpha\ ;\ n1-1\ ;n2-1} \Rightarrow R'H_0$

المبحث 4. اختبار الاستقلال والتجانس

اختبار التجانس
اختبار التعديل

1 اختبار التجانس

لنعد إلى اختبار النسبة، ونفترض أن لدينا عددا k من الخصائص المتنافية، نسبة تحقق كل منها في المجتمع p_i حيث $\sum p_i = 1$. نريد اختبار فرضية تساوي هذه النسب:

$$H_0 : p_i = p_{i0}\ \ (i = 1, 2, \ldots k) \leftrightarrow H_1 : p_i \neq p_{i0}$$

(الفرضية البديلة هي أن إحدى النسب النظرية p_{i0} على الأقل غير مساوية للقيمة الحقيقية.)

متغيرة القرار : لإنجاز الاختبار نستخرج عينة نحسب فيها عدد مرات تحقق الخصائص (n_i) . إذا تحققت الشروط

$n \geq 30$ ، $p_{i0} \geq 1$ وعلى الأقل في 80 % من الحالات $np_{i0} \geq 5$ نبرهن أن :

$$T = \sum_i \frac{(n_i - np_{i0})^2}{np_{i0}} \approx \chi^2_{k-1}$$

2 اختبار التعديل

تستخدم هذه الطريقة أيضا لاختبار تعديل توزيع معين بتوزيع آخر، وفي هذه الحالة نقارن بين تكرارات العينة (التكرارات الحقيقية) n_i وتكرارات افتراضية n_{i0} ، حيث تصاغ الفرضيات كما يلي :

$H_0: n_i = n_{i0}\ \ (i = 1, 2, \ldots k) \leftrightarrow H_1$: n_i على الأقل إحدى التكرارات النظرية n_{i0} غير مساوية للتكرار الحقيقي

$$T = \sum_i \frac{(n_i - n_{i0})^2}{n_{i0}} \approx \chi^2_{k-m-1}$$

m عدد من معالم من المجتمع المقدرة انطلاقا من بيانات العينة لتحديد التكرارات النظرية.

مثال: يتقدم إلى انتخابات معينة 3 مرشحين: أ، ب وج. نريد اختبار فرضية بمستوى معنوية 5 % حول شعبيتهم كما يلي:

$H_0 : p_1 = 0.4, p_2 = 0.35, p_3 = 25$

أجري استجواب ل 400 ناخب فكان توزع فئات المساندين على التوالي : 170، 135، 95 .

لدينا $n = 400 \geq 30$، الأعداد الافتراضية $np_{i0} = 160, 140, 100 \geq 1$ ، وأكثر من 80 % من $np_{i0} \geq 5$.

$$T = \sum_i \frac{(n_i - np_{i0})^2}{np_{i0}} = \frac{(170-160)^2}{160} + \frac{(135-140)^2}{140} + \frac{(95-100)^2}{100} = 1.05$$

$X^2_{2\ ;\ 0.95} = 5.99 > 1.05 \Rightarrow R'H_0$.