Statistiek
Hoorcollege 2, 17-11-2015

Probability theory

Waarschijnlijkheid houdt de link tussen populatie en steekproef in. Dit is essentieel voor het trekken van gevolgen. Veel beslissingen in de zakenwereld zijn gebaseerd op overtuigingen met betrekking tot de waarschijnlijkheid op een onzekere gebeurtenis. Voorbeelden hiervan zijn 'ik denk dat…', 'de kans is dat…' etc.

De meeste managers hebben moeite met het beoordelen van waarschijnlijkheid en vertrouwen op regels die ze zelf bedenken. Veel managers zien gelukstreffers dan ook vaak als uitkomst van hun eigen goede vaardigheden, terwijl dit helemaal niet het geval is.

Het rollen van een dobbelsteen is een random experiment. Het leidt tot verschillende mogelijke uitkomsten. De **sample space** S is een lijst van alle mogelijke uitkomsten. Een mogelijke uitkomst uit de sample space wordt een **simple event** genoemd.

Probability rules

- Complement rule: de kans dat gebeurtenis A niet gebeurt
  $P(A^c) = 1 - P(A)$
- Addition rule: de kans van de union van A en B (de kans dat A of B gebeurt)
  P(A of B) = P(A) + P(B) – P(A en B)
- Conditional probability: de kans dat gebeurtenis A zorgt voor een andere (mogelijk verbonden) gebeurtenis B
  $P\left(\frac{A}{B}\right) = \frac{P(A\ en\ B)}{P(B)}$
- Multiplication rule: de gezamenlijke kans op gebeurtenis A en B
  P(A en B) = P(B) x P($\frac{A}{B}$)
- For independent events
  P(A en B) = P(A) x P(B)

De wet van Baye

Er zijn veel implicaties in het maken van beslissingen in een bedrijf. Huidige of vroegere kansen statements kunnen gebruikt worden voor informatie.
Conditionele waarschijnlijkheid:
$P\left(\frac{A}{B}\right) = \frac{P(A\ en\ B)}{P(B)}$ is te herschrijven als P(A en B) = P(A) x P($\frac{B}{A}$)
De wet van Baye: $P\left(\frac{A}{B}\right) = \frac{P(A)x\ P(\frac{B}{A})}{P(B)}$

Kansbomen

Met behulp van populatie parameters kun je een kansboom opstellen. Voor voorbeeld: **<u>zie slide 10 en 11</u>**. Een uitkomst hiervan wordt een **random variable** genoemd, wat volgt uit een **binomial distributial**.

Een random variable is een functie dat een nummer aan elke mogelijke uitkomst toewijst. Voorbeeld: Als X het totaal is van 2 dobbelstenen is het een random variable dat de waardes 2 t/m 12 kan aannemen. Een **probability distribution** beschrijft de kans van de waardes die een random variable kan aannemen. De notatie om de kans dat random variable X kan hebben op waarde x gaat als volgt: P(X = x) of P(x). Voorwaarden voor distribution van een random variable zijn dat de P(x) tussen 0 en 1 moet liggen, en dat P(alle mogelijke x) = 1.

Binomial distribution

Binominale distributie is ontwikkeld door James Bernoulli in de 17^e eeuw . Het doel hiervan was het begrijpen van het concept onzekerheid. Dit vormt de basis van waarschijnlijkheids theorie. Het betreft een experiment dat bestaat uit een aantal n van onafhankelijke pogingen. Elke poging heeft 2 mogelijke uitkomsten: success (S) en failure (F). Elke poging heeft een kans op succes p, en een kans op falen van 1-p.

Als een experiment n keer herhaald wordt is de exacte volgorde van de uitkomsten vaak niet relevant. Je bent bij resultaten van series van success en failure namelijk op zoek naar het totaal aantal successen. De random variable hierbij is het aantal successen X, en de probability distribution is P(X=x).

Voor voorbeelden: **zie slide 17 t/m 26**

Constructing probability distributions

Een frequency table geeft waardes en vervolgens de frequenties aan. Voor voorbeelden: **zie slide 27 t/m 29**.

Calculating the mean of a discrete random variable

Je kunt een verwachte waarde formule opstellen zodat je niet alle observaties hoeft vast te leggen (**zie slide 30**).

Calculating the variabce of a discrete random variable

Voor formules: **zie slide 31.**

Voor voorbeelden: **zie slide 32 en 33**.

Covariance of a discrete random variable

Voor formules: **zie slide 34**

Verschillende types random variables:

- Discreet: dit kunnen een telbaar aantal waarden zijn (bijvoorbeeld het aantal besmette patiënten) en er liggen geen waarden tussen twee opeenvolgende waardes.
- Continu: de mogelijkheden zijn ontelbaar en er zijn ook ontelbaar veel waardes tussen twee waardes (bijvoorbeeld bij lengtes)

Over het algemeen zijn continue random variables gemakkelijker te gebruiken, ook als benadering van discrete random variables.

Voor voorbeelden: **zie slide 37 t/m 43**

Uniforme distributie

Uniforme distributie op het interval [a,b] heeft waarschijnlijkheids dichtheid functie $f\ (x) = \frac{1}{b-a'}$ waarbij x tussen a en b ligt. Voor voorbeeld: **zie slide 45 en 46.**

Normale distributie

Voor formule en voorbeeld: **zie slide 47 t/m 52.**