

Hertentamen: Statistisch Redeneren

docent: Rein van den Boomgaard

27 juni 2016

- Maak de opgaven in volgorde op uw antwoordvel
- Vermeld duidelijk naam en studienummer op elk antwoordvel
- Succes !!

1 Kansrekening I

Gegeven het kanstolletje met een bovenaanzicht zoals hier geschetst. Loodrecht op het aanzicht en door het midden zit een as. Nadat het tolletje uit is gedraaid zal het met gelijke kans op 1 van de 6 randen tot stilstand komen.

De stochast X heeft als uitkomst 1, 2 of 3 naargelang het tolletje tot stilstand komt in één van de genummerde gebieden.

1. Wat is de kans functie (probability mass function) $p_X(x) = P(X = x)$, voor alle mogelijke x , van de stochast X ?
2. Geef de *definitie* van de *verwachting* van een stochast X en bereken de verwachting voor de hierboven beschreven stochast.
3. Geef de definitie van de *variantie* van een stochast X en bereken deze voor de hierboven beschreven stochast.
4. Nu maken we een tolletje met precies dezelfde kleurenverdeling maar als informaticus zijnde beginnen we het nummeren met 0 en verhogen we steeds met 10. De uitkomsten zijn dan 0, 10 en 20. Noem dit stochast Y . Bereken $E(Y)$. En bereken $\text{Var}(Y)$.

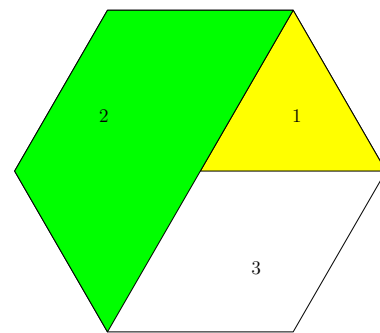


Figure 1: Een kanstol

2 Kansrekening II

Gegeven de stochasten X , Y en Z . Gegeven is dat $X \in \{1, 2, 3\}$, $Y \in \{1, 2\}$ en $Z \in \{1, 2\}$. De gezamenlijke kansdichtheid (joint distribution) is $p_{XYZ}(x, y, z) = P(X = x, Y = y, Z = z)$ en is geënumereerd in onderstaande tabel:

x	y	z	$p_{XYZ}(x,y,z)$
1	1	1	0.11
1	1	2	0.09
1	2	1	0.08
1	2	2	0.04
2	1	1	0.12
2	1	2	0.08
2	2	1	0.15
2	2	2	0.08
3	1	1	0.05
3	1	2	0.13
3	2	1	0.04
3	2	2	0.03

1. Bereken $P(Y = 2)$ en $P(X = 1, Z = 2)$
2. Bereken $P(Y = 2|X = 3, Z = 1)$.
3. Wanneer zijn twee events A en B onafhankelijk? Wanneer zijn twee discrete stochasten X en Y onafhankelijk?
4. Zijn X en Y zoals gedefinieerd in de bovenstaande gezamenlijke kanstabel onafhankelijke stochasten?

3 Continue Stochasten

1. De *verdelingsfunctie* voor continue stochast X is F_X . Geef de kans $P(a \leq X \leq b)$ in termen van de verdelingsfunctie.
2. Een verdeling wordt bepaald door zijn kansdichtheidsfunctie:

$$f_X(x) = \begin{cases} A\sqrt{x} & \text{als } 0 \leq x \leq 1 \\ 0 & \text{elders} \end{cases}$$

Bereken de waarde van A .

3. Bereken de verwachting voor de stochast X met kansdichtheidsverdeling zoals hierboven gegeven.
4. De verwachting van een normaal verdeelde stochast $Y \sim N(\mu, \sigma)$ is gelijk aan μ , d.w.z. $E(Y) = \mu$. Wat is nu de verwachting van $Z = X + Y$ als X de stochast is met kansdichtheidsfunctie gedefinieerd hierboven.

4 Naive Bayes Classifier

Beschrijf in woord en formule de Naive Bayes Classifier om een persoon waarvan het gewicht G , lengte L en schoenmaat S gegeven is te classificeren als man of vrouw. Van een N tal personen is bekend of het een man of vrouw is en wat de waarden van de drie kenmerken zijn.

5 Multivariate Stochasten

1. Geef de definitie van de covariantie matrix $\text{Cov}(\mathbf{X})$ van een vectoriele stochast $\mathbf{X}$ in termen van de covariantie van scalaire stochasten. Wat zijn de afmetingen van deze matrix?

2. Welke van de twee onderstaande stellingen is waar?

$$X \perp Y \rightarrow \text{Cov}(X, Y) = 0 \quad (1)$$

$$\text{Cov}(X, Y) = 0 \rightarrow X \perp Y \quad (2)$$

Hierin is $\rightarrow$ de implicatie pijl ($A \rightarrow B$: als A dan B). Maak uw antwoord ook aannemelijk (een bewijs is het mooiste...)

6 Principal Component Analysis

1. Leg uit in drie zinnen hoe je met PCA in staat bent om in veel praktijk situaties de dimensionaliteit van een probleem te verlagen.

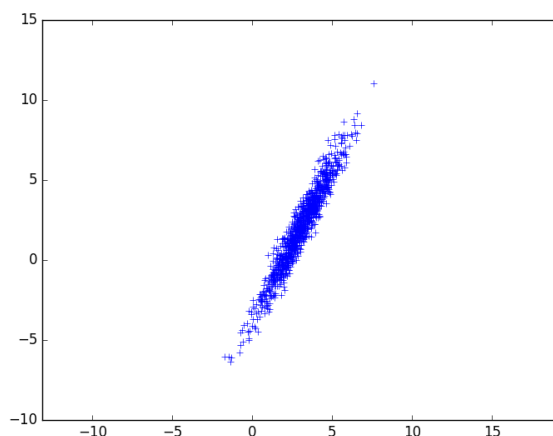


Figure 2: Scatterplot of 2D dataset

1. In bovenstaande figuur een scatterplot 1000 trekkingen uit een 2 dimensionale stochast $\mathbf{X}$. Gegeven is dat het gemiddelde is:

$$\bar{\mathbf{x}} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

en de schatter van de covariantiematrix

$$S = \begin{pmatrix} 1.88 & 3.56 \\ 3.56 & 7.22 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 0.1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T$$

Als we nu met een PCA de dimensionaliteit van 2 terugbrengen tot 1 kunnen we een willekeurige vector $\mathbf{x} = (a \ b)^T$ met één scalar representeren. Hoe bereken je die scalar? En hoe benader je $\mathbf{x}$ gegeven alleen de net berekende scalar.

7 Support Vector Machines

- Beschrijf de SVM in enkele zinnen.
- De *lineaire* SVM werkt niet in alle praktijk situaties. In welke gevallen is dat? En op welke manier helpt de "kernel trick" om in deze gevallen wel tot succesvolle classificatie te komen.

3. De standaard SVM is een twee klassen classifier. Hoe maak je het toch mogelijk om meerdere klassen uit elkaar te houden met behulp van deze SVM's.
4. De SVM met een radial basis functie kernel kent 2 parameters: C en γ . Hoe bepaal je de waarden van deze parameters om een optimale classifier te maken.

8 Clustering

1. Waarin verschilt clustering van classificatie?
2. Beschrijf het k-means clustering algoritme. Geef de nodige formules die een programmeur nodig heeft om het algoritme te implementeren.