

LÖSNINGAR till tentamen: Statistik och sannolikhetslära (LMA120)

Tid och plats: 8.30-12.30 den 24 augusti 2015

Hjälpmedel: Typgodkänd miniräknare, formelblad

Betygsgränser: 3: 12 poäng, 4: 18 poäng, 5: 24 poäng. Maximalt antal poäng: 30

Examinator: Petter Mostad

Telefonvakt: Malin Palö Forsström, telefonnummer: 0768-303635

1. Antag att man kastar en vanlig tärning och räknar antalet ögon man får. Visa att händelserna *tärningen visar ett udda antal ögon* och *tärningen visar ett eller två ögon* oberoende. (3p)

Lösning

$$P(\text{man får ett udda antal ögon}) = P(\{1, 3, 5\}) = 3/6 = 1/2$$

$$P(\text{man får ett eller två ögon}) = P(\{1, 2\}) = 2/6 = 1/3$$

och

$$P(\text{man får ett udda antal ögon} \text{ **och** man får ett eller två ögon}) = P(\text{man får ett öga}) = 1/6.$$

Eftersom att $1/6 = 1/2 \cdot 1/3$ så är de två händelserna oberoende.

2. Under det senaste året har ett antal tidningsartiklar hävdat att cirka 60% av de som tror att de är laktosintoleranta i själva verket inte är det. Totalt är 15% av befolkningen laktosintoleranta. Av de som är laktosintoleranta antas de flesta, 80%, veta om det.
 - (a) Hur stor är sannolikheten att en slumpvist vald person tror att den är laktosintolerant? (2p)
 - (b) Att ha en anställd med laktosintolerans ger företaget Alfreds bygg merkostnad en merkostnad på cirka 200 kr per år. Ett hemmatest för laktosintolerans kostar 399 kronor på apoteket. Alfred funderar därför på att ge nyanställda som uppger att de är laktosintoleranta ett test från Apoteket, för att eventuellt slippa merkostnaden om de skulle visa sig inte vara laktosintoleranta. Om en anställd stannar hos företaget i tre år, hur stor är den förväntade vinsten med att köpa ett sådant test? (3p)

Lösning

- (a) Låt T vara händelsen att en slumpvist vald person tror att den är laktosintolerant, och låt L vara händelsen att en slumpvist vald person är laktosintolerant. Enligt uppgiften är $P(L | T) = 0.4$, $P(L) = 0.15$ och $P(T | L) = 0.8$. Vi ska räkna ut $P(T)$ och använder Bayes sats:

$$P(T) = \frac{P(T | L)P(L)}{P(L | T)} = \frac{0.8 \cdot 0.15}{0.4} = 0.3$$

dvs. 30% av befolkningen tror att de är laktosintoleranta.

- (b) Om man inte testar den anställda blir den förväntade kostnaden för tre år $3 \cdot 300 = 900$ kr enligt uppgiften. Om arbetsgivaren köper ett test blir den förväntade kostanden 399 kr om personen i själva verket inte är laktosintolerant, och $399 + 900 = 1299$ kr om den är det. Enligt uppgiften är sannolikheten att den nyanställda är laktosintolerant 40%. Den förväntade kostnaden om mna köper in ett test blir därför

$$1299 \cdot 0.4 + 399 \cdot 0.6 = 759.$$

Arbetsgivaren kan alltså spara i snitt $900 - 759 = 141$ kronor per nyanställd på att genomföra ett sådant test.

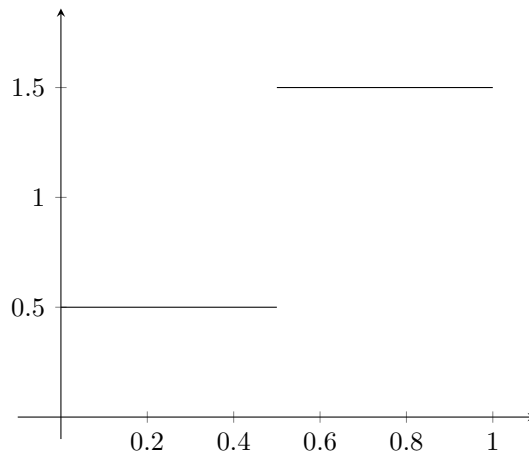
3. Antag att X är likformigt fördelad på intervallet mellan 0 och 1, och att

$$Y = \begin{cases} 2X & \text{om } X \leq 0.5 \\ X & \text{annars} \end{cases}$$

- (a) Rita täthetsfunktionen till Y (1p)
(b) Beräkna $P(Y) > 0.25$ (1p)
(c) Beräkna $\mathbb{E}[Y]$ (2p)
(d) Beräkna $\text{Var}(Y)$ (2p)

Lösning

- (a) Se nedan.



- (b) Från figuren ovan ses direkt att $P(Y > 0.25) = 7/8$. Ett annat sätt att räkna ut det på är:

$$P(Y > 0.25) = 1 - P(Y \leq 0.25) = 1 - P(X \leq 0.125) = 1 - 1/8 = 7/8.$$

- (c) Låt f vara täthetsfunktionen till Y . Då är

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^1 x f(x) dx = \int_0^{0.5} x \cdot 0.5 dx + \int_{0.5}^1 x \cdot 1.5 dx \\ &= 0.5 \left[x^2/2 \right]_0^{0.5} + 1.5 \left[x^2/2 \right]_{0.5}^1 = 0.5 \cdot (0.5^2/2 - 0) + 1.5 \cdot (1^2/2 - 0.5^2/2) \\ &= 0.6250 \end{aligned}$$

(d) Vi räknar först ut $\mathbb{E}[Y^2]$:

$$\begin{aligned}\mathbb{E}[Y^2] &= \int_0^1 x^2 f(x) dx = \int_0^{0.5} x^2 \cdot 0.5 dx + \int_{0.5}^1 x^2 \cdot 1.5 dx \\ &= 0.5 \left[x^3/3 \right]_0^{0.5} + 1.5 \left[x^3/3 \right]_{0.5}^1 = 0.5 \cdot (0.5^3/3 - 0) + 1.5 \cdot (1^3/3 - 0.5^3/3) \\ &= 0.4583\end{aligned}$$

Vi får nu

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = 0.4583 - 0.6250^2 = 0.0677$$

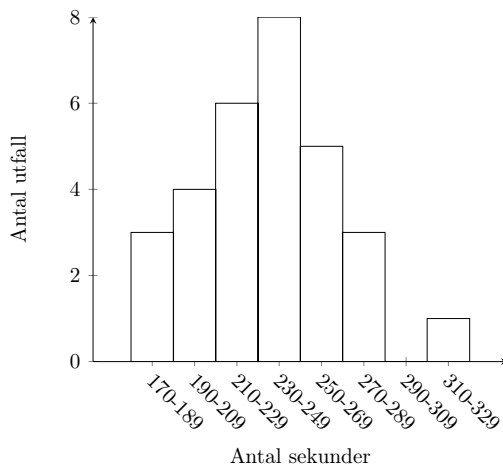
4. Listan nedan visar tidsåtgången i sekunder det tagit för ett visst typ av klister att härda.

327	265	281	208	226	232	273	232	261	178
229	215	193	255	248	241	200	274	251	231
241	232	187	231	215	211	205	224	180	269

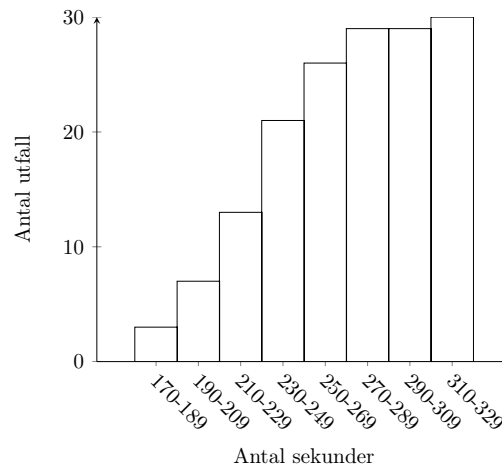
- (a) Rita ett stapeldiagram som sammanfattar datan. Rita också en empiriska fördelningsfunktion för datan. (2p)
- (b) Vilket av de fördelningar (med vilka parametrar) som vi gått igenom i kursen tycker du passar datan bäst? Finns det någonting som gör att den fördelning du valt inte kan stämma helt? För full poäng krävs full motivering. (2p)

Lösning

(a) Se figur 1 och 2.



Figur 1: Stapeldiagram



Figur 2: Empirisk fördelningsfunktion

- (b) Eftersom att tiderna anges i sekunder är det svårt att enbart från diagrammet säga om datan kommer från en diskret eller kontinuerlig fördelning, men eftersom att vi vet att den anger härdningstider, så borde datan komma från en kontinuerlig fördelning. Datat skulle kunna komma från en normalfördelning. Argument för att det skulle kunna vara så är att stapeldiagrammet har ungefär en sådan form. Vi kan skatta parametrarna till en normalfördelning med stickprovsmedelvärdet och stickprovsvariansen: $\bar{x} = 233.83$ och $s = 33.22$.

5. Den 17 juni i år skrev tidningen GP följande:

Miljöpartiet får sitt bästa stöd hittills i år och ökar med 2,3 procentenheter, enligt Sifos nya mätning. Samtidigt minskar stödet för regeringspartnern S med 2,4 procentenheter jämfört med i mars, ett tapp som dock inte är statistiskt säkerställt.

Förklara vad det betyder att tappet inte är statistiskt säkerställt. Hur skulle de som har gjort undersökningen ha räknat för att komma fram till det? (6p)

Lösning För att kontrollera om en sådan förändring är statistiskt säkerställd så tänker man sig att varje person i stickprovet representeras en Bernoullivariabel X_i som är 1 om en person anger att han/hon ska rösta på S och 0 annars. Då är X_i en slumpvariabel som är Bernoullifördelad med parameter p , där p är andelen i befolkningen som anger att de ska rösta på S. p är dessutom medelvärdet för X_i . När man har samlat in data har man gjort ett konfidensintervall för medelvärdet, dvs. för p , med någon konfidensgrad. Om det intervallet inte innehåller det värde man hade sedan tidigare, exempelvis från en tidigare undersökning, så säger man att skillnaden eller förändringen är statistiskt säkerställd. I undersökningen som GP skrev om så låg alltså det gamla värdet i det konfidensintervall man fick från den nya undersökningen, så det går inte att säga om det verkligen är fler som har tänkt rösta på S, eller om skillnaden i resultat bara beror på slumpen.

6. Om man bygger ett hus nära en stor väg kan man behöva sätta upp bullerplank om den genomsnittliga ljudnivån inomhus per dygn överstiger riktvärdet 30 dBA. För att kontrollera om detta var fallet gjordes följande mätningar i ett hus under ett dygn (i dBA).

25.9	28.3	29.7	26.4	29.2	29.0	27.7	28.4
31.0	26.7	29.0	28.6	28.8	28.4	28.7	28.1
26.4	27.7	28.7	28.7	29.1	31.1	26.4	29.2

- (a) Antag att stickprovet kommer från en normalfördelning och beräkna ett 99%-igt konfidensintervall för den genomsnittliga bullernivån per dygn. Du kan använda att $\sum x_i = 681.2$ och $\sum x_i^2 = 19375.64$. (2p)
- (b) I (a)-uppgiften antog vi att datan var normalfördelad. Testa om det verkar vara ett rimligt antagande med hjälp av ett hypotestest. För full poäng krävs att en korrekt slutsats dras. (4p)

Lösning

- (a) Eftersom att vi har 24 mätpunkter så får vi 23 frihetsgrader. Enligt tabell vet vi att ett 99%-igt konfidensintervall ges av

$$[\bar{x} - 2.81s/\sqrt{n}, \bar{x} + 2.81s/\sqrt{n}],$$

där

$$\bar{x} = \frac{\sum x_i}{24} = 28.38$$

och

$$s = \sqrt{\frac{\sum x_i^2 - (\sum x_i)^2/n}{n-1}} = 1.334.$$

Konfidensintervallet blir därför

$$[27.5984, 29.1616].$$

(b) Vi sammanfattar först datan i en tabell:

i	Ljudnivå (dBA)	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
1	< 26.5	5			
2	$26.5 - 26.99$	6			
3	$27.0 - 28.99$	5			
4	≥ 29	8			

Här har vi slagit samman intervall så att $O_i \geq 5$ för alla intervall. För att kunna fylla i resten av tabellen behöver vi använda hypotesen, nämligen att datan är normalfördelad, och gissningarna på parametrarna som ges av $\bar{x}$ och s . Vi räknade ut dessa i (a)-uppgiften, så vi använder fördelningen $N(28.38, 1.334)$ and antar att $X \sim N(28.38, 1.334)$ för att räkna ut kolumnen för E_i :

$$\begin{aligned} E_1 &= 28 \cdot P(X < 26.5) = 28 \cdot \Phi\left(\frac{26.5 - 28.38}{1.334}\right) \\ &= 28 \cdot \Phi(-1.4093) = 28 \cdot (1 - \Phi(1.4093)) \\ &= 28 \cdot (1 - 0.9206) = 28 \cdot 0.0794 = 2.2225 \end{aligned}$$

$$\begin{aligned} E_2 &= 28 \cdot P(26.5 \leq X < 27) = 28 \cdot (P(X < 27) - P(X < 26.5)) \\ &= 28 \cdot \left(\Phi\left(\frac{27 - 28.38}{1.334}\right) - 0.0794\right) = 28 \cdot (\Phi(-1.0345) - 0.0794) \\ &= 28 \cdot (1 - \Phi(1.0345) - 0.0794) = 28 \cdot (1 - 0.8495 - 0.0794) \\ &= 28 \cdot (0.1505 - 0.0794) = 28 \cdot 0.0711 = 1.99 \end{aligned}$$

$$\begin{aligned} E_3 &= 28 \cdot P(27 \leq X < 29) = 28 \cdot (P(X < 29) - P(X < 27)) \\ &= 28 \cdot \left(\Phi\left(\frac{29 - 28.38}{1.334}\right) - 0.1505\right) = 28 \cdot (\Phi(0.4648) - 0.1505) \\ &= 28 \cdot (0.6790 - 0.1505) = 28 \cdot 0.5285 = 14.80 \end{aligned}$$

$$\begin{aligned} E_4 &= 28 \cdot P(X \geq 29) = 28 \cdot (1 - P(X < 29)) \\ &= 28 \cdot \left(1 - \Phi\left(\frac{29 - 28.38}{1.334}\right)\right) \\ &= 28 \cdot (1 - \Phi(0.4648)) = 28 \cdot (1 - 0.6790) = 9.00 \end{aligned}$$

Om vi sätter in det här i tabellen får vi:

i	Ljudnivå (dBA)	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
1	≤ 26.4	5	2.22	2.78	3.88
2	$26.5 - 26.9$	6	1.99	4.01	8.08
3	$27.0 - 28.9$	5	14.80	- 9.80	6.49
4	≥ 29	8	9.00	-1.00	0.11

Vi får alltså statistikan

$$\sum_i (O_i - E_i)^2 / E_i = 3.88 + 8.08 + 6.49 + 0.11 = 18.56$$

Vi har $4 - 1 = 3$ frihetsgrader vilket gör att vi hade förkastat vår hypotes även om signifikansnivån hade varit så låg som 0.005.