STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

# TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR I
2023-03-24
LÖSNINGAR

**Uppgift 1.** (20 poäng)

En övningslärare i statistik bakar semlor för att bjuda sina kompisar på institutionen. Hon gör några med mandelmassa och några utan mandelmassa. Hon bedömer att antal av vardera sorten som äts upp ges av den simultana sannolikhetsfunktionen

$$p\left(y_1, y_2\right) = \frac{y_1 + y_2 + 1}{64}, \;\; y_1 = 0, 1, 2, 3 \text{ och } y_2 = 0, 1, 2, 3$$

där $y_1$ är antal semlor med mandelmassa och $y_2$ är antal semlor utan mandelmassa.

a) Bestäm marginalsannolikheterna för $Y_1$ och $Y_2$.

b) Undersök om $Y_1$ och $Y_2$ är stokastiskt oberoende.

c) Bestäm sannolikhetsfördelningen för totala antalet antal semlor som äts upp, alltså sannolikhetsfördelningen för $S = Y_1 + Y_2$.

d) Bestäm förväntat antal semlor som äts upp, dvs. $E\left(S\right)$.

**Lösning:**
a) Simultana sannolikheterna beräknas enklas i en tabell. Marginalsannolikheterna fås genom att summera rader respektive kolumner.

| | | $Y_2$ | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | $p_2\left(y_2\right)$ |
| $Y_1$ | 0 | $\frac{1}{64}$ | $\frac{2}{64}$ | $\frac{3}{64}$ | $\frac{4}{64}$ | $\frac{10}{64}$ |
| | 1 | $\frac{2}{64}$ | $\frac{3}{64}$ | $\frac{4}{64}$ | $\frac{5}{64}$ | $\frac{14}{64}$ |
| | 2 | $\frac{3}{64}$ | $\frac{4}{64}$ | $\frac{5}{64}$ | $\frac{6}{64}$ | $\frac{18}{64}$ |
| | 3 | $\frac{4}{64}$ | $\frac{5}{64}$ | $\frac{6}{64}$ | $\frac{7}{64}$ | $\frac{22}{64}$ |
| | $p_1\left(y_1\right)$ | $\frac{10}{64}$ | $\frac{14}{64}$ | $\frac{18}{64}$ | $\frac{22}{64}$ | 1 |

b) Om $Y_1$ och $Y_2$ är stokastiskt oberoende gäller det att $p\left(y_1, y_2\right) = p_1\left(y_1\right) p_2\left(y_2\right)$ för alla $y_1$ och $y_2$. Men $p\left(0, 0\right) = \frac{1}{64}$ medan $p_1\left(0\right) p_2\left(0\right) = \frac{10}{64}\frac{10}{64} = \frac{100}{4096} = \frac{25}{1024}$, dvs. $p\left(0, 0\right) \neq p_1\left(0\right) p_2\left(0\right)$ och vi drar slutsatsen att $Y_1$ och $Y_2$ är stokastiskt beroende.
c) Man får sannolikheter för $S$ ur tabellen i uppgift b, t.ex. $p_S\left(0\right) = P\left(Y_1 = 0, Y_2 = 0\right) = p\left(0, 0\right) = \frac{1}{64}$, $p_S\left(1\right) = P\left(\left(Y_1 = 1, Y_2 = 0\right) \cup \left(Y_1 = 0, Y_2 = 1\right)\right) = p\left(1, 0\right) + p\left(0, 1\right) = \frac{2}{64} + \frac{2}{64} = \frac{4}{64}$. Sannolikhetsfunktionen för S sammanfattas i nedanstående tabell

| $s$ | $p_S(s)$ | $sp_S(s)$ |
|---|---|---|
| 0 | $\frac{1}{64}$ | 0 |
| 1 | $\frac{4}{64}$ | $\frac{4}{64}$ |
| 2 | $\frac{9}{64}$ | $\frac{18}{64}$ |
| 3 | $\frac{16}{64}$ | $\frac{48}{64}$ |
| 4 | $\frac{15}{64}$ | $\frac{60}{64}$ |
| 5 | $\frac{12}{64}$ | $\frac{60}{64}$ |
| 6 | $\frac{7}{64}$ | $\frac{42}{64}$ |
| Summa | 1 | $\frac{232}{64}$ |

d) Beräkning av $E(S) = \sum sp_S(s)$ framgår av tabellen ovan. Förväntat antal semlor som äts upp är $\frac{232}{64} = \frac{29}{8} = 3\frac{5}{8}$.

**Uppgift 2.** (20 poäng)

Den stokastiska variabeln $Y$ har sannolikhetsfunktionen

$$p(y) = \begin{cases} \frac{y}{15} \text{ om } y = 1, 2, \ldots, c \\ 0 \text{ för övrigt} \end{cases}$$

a) Bestäm konstanten $c$ så att $p(y)$ är en sannolikhetsfunktion.

b) Bestäm fördelningsfunktionen $F(y)$. Låt $F(y)$ vara en funktion av $c$ om du inte har löst uppgift a).

c) Beräkna $P(2 \leq Y \leq 4)$.

**Lösning:**
Vi vet att

$$1 = \sum_{y=1}^{c} p(y) = \frac{1}{15} + \frac{2}{15} + \ldots + \frac{c}{15} = \frac{1 + 2 + \ldots + c}{15} = \frac{c(c+1)/2}{15}$$

vilket ger att

$$c(c+1) = 30$$

dvs. $c = 5$. Resultatet kan även får genom att prova sig fram med $c = 1$, $c = 2$ osv. tills summan $\sum_{y=1}^{c} p(y)$ blir 1.

b) Fördelningsfunktionen fås genom tabellen

| $y$ | $p(y)$ | $F(y)$ |
|---|---|---|
| 1 | 1/12 | 1/12 |
| 2 | 2/12 | 3/12 |
| 3 | 3/12 | 6/12 |
| 4 | 4/12 | 10/12 |
| 5 | 5/12 | 1 |

c) $P(2 \leq Y \leq 4) = P(1 < Y \leq 4) = F(4) - F(1) = \frac{10}{12} - \frac{1}{12} = \frac{9}{12}$ alternativt kan man beräkna $P(2 \leq Y \leq 4) = p(2) + p(3) + p(4) = \frac{2}{12} + \frac{3}{12} + \frac{4}{12} = \frac{9}{12}$.

**Uppgift 3.** (20 poäng)

Snödjupet i Stockholm en slumpmässigt vald dag i januari kan beskrivas med en exponentialfördelad stokastisk variabel med väntevärdet $\beta = 8$ cm. Om snödjupet överstiger 10 cm utfärdas en röd varning och om snödjupet är mellan 5 och 10 cm utfärdas en gul varning,

a) Bestäm sannolikheterna för att röd varning, gul varning respektive ingen varning utfärdas en slumpmässigt vald dag i januari.

b) Tio januaridagar väljs slumpmässigt ut och låt $X_1$ vara antal dagar utan varning, $X_2$ vara antal dagar med gul varning och $X_3$ vara antal dagar med röd varning. Vad heter den simultana fördelningen för $X_1$, $X_2$ och $X_3$?

c) Beräkna förväntade värden för $X_1$, $X_2$ och $X_3$.

d) Vad är sannolikheten att ingen röd varning utfärdas under de 10 dagarna?

**Lösning:**
a) Låt $Y$ vara snödjupet. Då är $Y \sim \exp(\beta = 8)$.
Vi söker $P(\text{röd varning}) = P(Y \geq 10) = \int_{10}^{\infty} \frac{1}{8} e^{-y/8} dy = \left[-e^{-y/8}\right]_{10}^{\infty} = 0 - \left(-e^{-10/8}\right) = e^{-1,25} \approx 0,2865$
$P(\text{gul varning}) = P(5 \leq Y \leq 10) = \int_{5}^{10} \frac{1}{8} e^{-y/8} dy = \left[-e^{-y/8}\right]_{5}^{10} = e^{-0,625} - e^{-1,25} \approx 0,24876$
$P(\text{ingen varning}) = P(Y \leq 5) = \int_{0}^{5} \frac{1}{8} e^{-y/8} dy = \left[-e^{-y/8}\right]_{0}^{5} = 1 - e^{-0,625} \approx 0,46474$
Koll att sannolikheterna summerar till 1: $0,2865 + 0,24876 + 0,46474 = 1,0$. Stämmer!

b) Multinomial fördelning (eller trinomial fördelning).

c) $E(X_1) = 10 \cdot 0,46474 = 4,6474$, $E(X_2) = 10 \cdot 0,24876 = 2,4876$, $E(X_3) = 10 \cdot 0,2865 = 2,865$.

d) $P(\text{ingen röd varning}) = (1 - 0.2865)^{10} = 0.7135^{10} \approx 0,03419\,3$

**Uppgift 4.** (20 poäng)

Ett företag köper in partier av komponenter från en leverantör till sin tillverkning. När ett parti kommer in till företaget testas kvaliteten hos partiet genom att man tar ett slumpmässigt urval komponenter. Om urvalet innehåller för många felaktiga enheter återsänds partiet till leverantören. Partiet är så stort att tester av enskilda komponenter kan anses vara stokastiskt oberoende av varandra.

a) Antag att man tar ett slumpmässigt urval om 10 komponenter och återsänder partiet om 3 eller fler komponenter inte fungerar. Bestäm förväntat antal felaktiga komponenter i urvalet och sannolikheten att partiet accepteras om partiet innehåller 10 procent felaktiga komponenter.

b) Antag att man tar ett slumpmässigt urval om 20 komponenter och återsänder partiet om 6 eller fler komponenter inte fungerar. Bestäm förväntat antal felaktiga komponenter i urvalet och sannolikheten att partiet accepteras om partiet innehåller 10 procent felaktiga komponenter.

**Lösning:**
Låt $Y$ vara antal felaktiga komponenter i ett urval. Eftersom va kan anta oberoende mellan komponenterna är $Y\,(n, p)$, där n är urvalsstorleken och $p$ är andelen felaktiga komponenter i partiet.

a) $n = 10$ och $p = 0.10$. Vi får då $E\,(Y) = 10 \cdot 0,1 = 1$ och $P\,(\text{partiet accepteras}) = P\,(Y \leq 2) = 0,9298$

b) $n = 20$ och $p = 0.10$. Vi får då $E\,(Y) = 20 \cdot 0,1 = 2$ och $P\,(\text{partiet accepteras}) = P\,(Y \leq 4) = 0,9568$

**Uppgift 5.** (20 poäng)

Låt $Y$ vara en normalfördelad stokastisk variabel med väntevärdet $\mu$ och variansen $\sigma^2$. Bilda en ny variabel genom standardiseringen

$$Z = \frac{Y - \mu}{\sigma}.$$

a) Bestäm täthetsfunktionen för $Z$.
b) Vad kallas fördelningen för $Z$?

**Lösning:**
a) $z = h(y) = \frac{y-\mu}{\sigma}$ är en monoton funktion av $y$ så vi kan använda transformationsmetoden. Notera att $Z$ kan anta alla reella tal. Utfallsrummet för $U$ är alltså intervallet $(-\infty, \infty)$. Vi ser att $\sigma z = y - \mu$ så att $y = h^{-1}(z) = \sigma z + \mu$. Då är $\frac{dh^{-1}}{dz} = \sigma$ och vi får att täthetsfunktionen för $Z$ är $f_Z(z) = f_Y\left(h^{-1}(z)\right)\left|\frac{dh^{-1}}{dz}\right|$.
Från tabellsamlingen vet vi att $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$. Detta ger nu

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\right)\sigma = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right).$$

b) Täthetsfunktionen i uppgift a är täthetsfunktionen för en standard normal, $N(0,1)$, fördelning.