STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR I
2023-02-13
LÖSNINGAR

Uppgift 1. (20 poäng)

En övningslärare i statistik gillar semlor och köper därför fyra stycken semlor på ett bageri för att bjuda sina kompisar. Hon anser att en semla är lagom stor om den väger mellan 120 och 140 gram. Antag att semlornas vikt är normalfördelad med väntevärdet 130 gram och standardavvikelsen 10 gram.

a) Beräkna sannolikheten att en slumpmässigt vald semla uppfyller övningslärarens krav, dvs. har en vikt mellan 120 och 140 gram?

b) Övningsläraren väger semlorna en efter en tills hon hittar fyra stycken med lämplig vikt. Vad är förväntat antal semlor hon behöver väga?

c) Vad är sannolikheten att övningsläraren behöver väga fler semlor än förväntat för att hitta fyra stycken med lämplig vikt?

Lösning:
a) Låt Y vara vikten hos en semla. Då är $Y \sim N\left(\mu = 130, \sigma^2 = 10^2\right)$. Vi söker $P\left(120 \leq Y \leq 140\right) = P\left(\frac{120-130}{10} \leq \frac{Y-\mu}{\sigma} \leq \frac{140-130}{10}\right) = P\left(-1 \leq Z \leq 1\right) = \Phi\left(1\right) - \Phi\left(-1\right) \approx 0.8413 - 0.1587 = 0.6826$
b) Låt U vara antal semlor hon behöver väga. Då är $U \sim Negbin\left(r = 4, p = 0.6826\right)$. Då följer det att $E\left(U\right) = \frac{r}{p} = \frac{4}{0.6826} = 5.8599$
c) Vi söker $P\left(U > E\left(U\right)\right) = P\left(U \geq 6\right) = 1 - P\left(U \leq 5\right) = 1 - \left(P\left(U = 4\right) + P\left(U = 5\right)\right) = 1 - \left(\binom{3}{3} 0.6826^4 \left(1 - 0.6826\right)^0 + \binom{4}{3} 0.6826^4 \left(1 - 0.6826\right)^1\right) = 1 - \left(0.2171 + 0.2753\right) = 0.5076$

Uppgift 2. (20 poäng)

Den stokastiska variabeln Y har täthetsfunktionen

$$f(y) = \begin{cases} \frac{y(1-y)}{2} \text{ om } 0 \le y \le 1 \\ \frac{1}{12} \text{ om } 1 < y \le c \\ 0 \text{ för övrigt} \end{cases}$$

a) Bestäm konstanten c så att $f(y)$ är en täthetsfunktion.

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

c) Beräkna $P(0.5 \le Y \le 1.5)$.

Lösning:
Vi vet att

$$1 = \int_{-\infty}^{\infty} f(y)\,dy = \int_0^1 \frac{y(1-y)}{2}dy + \int_1^c \frac{1}{12}dy = \left[\frac{y^2/2 - y^3/3}{2}\right]_0^1 + \left[\frac{y}{12}\right]_1^c = \frac{1}{12} - 0 + \frac{c}{12} - \frac{1}{12} = \frac{c}{12}$$

vilket ger att

$$c = 12$$

b) För $y < 0$ är $F(y) = 0$.
I intervallet $[0,\ 1]$ är $F(y) = \int_0^y \frac{t(1-t)}{2}dt = \left[\frac{t^2/2 - t^3/3}{2}\right]_0^y = \frac{y^2/2 - y^3/3}{2} = \frac{1}{4}y^2 - \frac{1}{6}y^3$
I intervallet $[1, 12]$ är $F(y) = \frac{1}{12} + \int_1^y \frac{1}{12}dt = \frac{1}{12} + \left[\frac{t}{12}\right]_1^y = \frac{1}{12} + \frac{y}{12} - \frac{1}{12} = \frac{y}{12}$
För $y > 12$ är $F(y) = 1$

c) $P(0.5 \le Y \le 1.5) = F(1.5) - F(0.5) = \frac{1.5}{12} - \left(\frac{1}{4}0.5^2 - \frac{1}{6}0.5^3\right) = \frac{1}{8} - \left(\frac{1}{16} - \frac{1}{24}\right)$
$= \frac{18}{144} - \frac{9}{144} + \frac{3}{144} = \frac{12}{144} = \frac{1}{12} \approx 0.0833$

Uppgift 3. (20 poäng)

Poängen på en tentamen i 'Statistisk teori med tillämpningar' kan anses vara normalfördelad med väntevärdet $\mu = 70$ och variansen σ^2. Man vet också att 70% av de som tenterar blir godkända, dvs. får en poäng som är minst 50.

a) Bestäm variansen σ^2 i fördelningen för betygspoängen.

b) Tentamenspoängen klassificeras i tre klasser, U (underkänd) om poängen är under 50, G (godkänd) om poängen är mellan 50 och 80 och V (väl godkänd) om poängen överstiger 80. Beräkna sannolikheten att en slumpmässigt vald student får en poäng i klasserna U, G respektive V. Använd σ från a-uppgiften (eller hitta på ett lämpligt värde om du inte har löst den).

c) Antag att 20 studenter deltar i en tentamen och låt X_1 vara antal underkända (U), X_2 vara antal godkända (G) och X_3 vara antal väl godkända. Vad heter den simultana fördelningen för X_1, X_2 och X_3?

d) Beräkna förväntade värden för X_1, X_2 och X_3.

Lösning:
a) Låt Y vara tentamenspoängen. Då är $Y \sim N\left(\mu = 70, \sigma^2\right)$.
Vi vet att $P\left(Y \geq 50\right) = P\left(\frac{Y-\mu}{\sigma} \geq \frac{50-70}{\sigma}\right) = 1 - P\left(Z \leq -\frac{20}{\sigma}\right) = 0.7$
Eftersom $P\left(Z \leq -0.52\right) = 0.3$ följer det att $-\frac{20}{\sigma} = -0.52$, dvs. $\sigma = \frac{20}{0.52} = 38.462$ och alltså är variansen $\sigma^2 = 38.462^2 = 1479.3$.

b) Vi vet redan från uppgiftsformuleringen att $P\left(U\right) = 0.30$. Vi får att $P\left(V\right) = P\left(Y > 80\right) = 1 - P\left(\frac{Y-\mu}{\sigma} \leq \frac{80-70}{38.462}\right) = 1 - \Phi\left(0.26\right) = 1 - 0.6026 = 0.3974$. $P\left(G\right) = 1 - \left(0.30 + 0.3974\right) = 0.3026$

c) Multinomial fördelning (eller trinomial fördelning).

d) $E\left(X_1\right) = 20 \cdot 0.30 = 6$, $E\left(X_2\right) = 20 \cdot 0.3026 = 6.052$, $E\left(X_3\right) = 20 \cdot 0.3974 = 7.948$.

Uppgift 4. (20 poäng)

I en bokhandel säljer man de två populära böckerna 'Statistikens grunder' och 'Mathematical Statistics with Applications'. De simultana sannolikheterna för antal exemplar av vardera boken som en slumpmässigt vald kund köper ges av

$$p(y_1, y_2) = \frac{(y_1 - y_2)^2 + y_1}{44}, \; y_1 = 0, 1, 2, 3 \text{ och } y_2 = 0, 1, 2,$$

där y_1 är antal exemplar av 'Statistikens grunder' och y_2 är antal exemplar av 'Mathematical Statistics with Applications'.

a) Bestäm marginalsannolikheterna för vardera boken.

b) Är försäljningen av de båda böckerna stokastiskt oberoende av varandra?

c) Bestäm den betingade fördelningen för antal sålda exemplar av 'Statistikens grunder' givet att en kund köper ett exemplar av 'Mathematical Statistics with Applications'.

Lösning:
a) Vi börjar med att skriva upp de simultana sannolikheterna i en tabell. Marginalsannolikheterna får om man summerar rader respektive kolumner i tabellen.

		y_1				
		0	1	2	3	$p_{y_2}(y_2)$
	0	0	2/44	6/44	12/44	20/44
y_2	1	1/44	1/44	3/44	7/44	12/44
	2	4/44	2/44	2/22	4/44	12/44
	$p_{y_1}(y_1)$	5/44	5/44	11/44	23/44	1.00

b) Försäljningen av böckerna är stokastiskt oberoende om $p(y_1, y_2) = p_{y_1}(y_1)\, p_{y_2}(y_2)$ för alla värden på y_1 och y_2. I tabellen ser vi att $p_{y_1}(0)\, p_{y_2}(0) = \frac{5}{44}\frac{20}{44} \neq 0 = p(0, 0)$, dvs. försäljningen av de båda böckerna är stokastiskt beroende.
c) Betingade sannolikheter ges av $p_{1|2}(y_1 \mid y_2 = 1) = p(y_1, 1) / p_{y_2}(1)$. Resultaten redovisas i följande tabell:

y_1	$p_{1\|2}(y_1 \mid y_2 = 1)$
0	(1/44)/(12/44)=1/12
1	(1/44)/(12/44)=1/12
2	(3/44)/(12/44)=3/12
3	(7/44)/(12/44)=7/12
Summa	12/12=1

Uppgift 5. (20 poäng)

En grupp studenter på kursen 'Statistisk teori med tillämpningar' har en inlämningsuppgift som handlar om överlevnadsanalys. En av deluppgifterna består i att simulera livslängden för en viss typ av komponenter. Sannolikhetsfördelningen för livstiden ska vara en s.k. Weibullfördelning och man vet att man kan generera Weibullfördelade slumptal i R genom att bilda

$$U = \lambda \left(-\ln\left(1 - Y\right)\right)^{1/\kappa},$$

där Y är en likformigt fördelad stokastisk variabel i intervallet $[0,\ 1]$ och $\lambda > 0$ och $\kappa > 0$ är parametrar i Weibullfördelningen.

a) Bestäm täthetsfunktionen för en Weibullfördelad stokastisk variabel, dvs. bestäm täthetsfunktionen för U.

b) Om parametern $\kappa = 1$ i Weibullfördelningen får vi en fördelning vi känner väl till. Vilken är fördelningen?

Lösning:

a) $u = h(y) = \lambda\left(-\ln\left(1-y\right)\right)^{1/\kappa}$ är en monoton funktion av y så vi kan använda transformationsmetoden. Notera att U kan variera mellan 0 (om $Y = 1$) och ∞ (om $Y = 0$). Utfallsrummet för U är alltså intervallet $[0, \infty)$.

Vi ser att $u^\kappa = \lambda^\kappa\left(-\ln\left(1-y\right)\right)$, $\left(\frac{u}{\lambda}\right)^\kappa = \left(-\ln\left(1-y\right)\right)$, $(1-y) = \exp\left(-\left(\frac{u}{\lambda}\right)^\kappa\right)$ och $y = h^{-1}(u) = 1 - \exp\left(-\left(\frac{u}{\lambda}\right)^\kappa\right)$. Då är $\frac{dh^{-1}}{du} = -\exp\left(-\left(\frac{u}{\lambda}\right)^\kappa\right)\left(-\left(\frac{u}{\lambda}\right)^{\kappa-1}\right)\kappa\frac{1}{\lambda} = \frac{\kappa}{\lambda}\left(\frac{u}{\lambda}\right)^{\kappa-1}\exp\left(-\left(\frac{u}{\lambda}\right)^\kappa\right)$ och vi får att täthetsfunktionen för U är $f_U(u) = f\left(h^{-1}(u)\right)\left|\frac{dh^{-1}}{du}\right| = 1\left|\frac{\kappa}{\lambda}\left(\frac{u}{\lambda}\right)^{\kappa-1}\exp\left(-\left(\frac{u}{\lambda}\right)^\kappa\right)\right| = \frac{\kappa}{\lambda}\left(\frac{u}{\lambda}\right)^{\kappa-1}\exp\left(-\left(\frac{u}{\lambda}\right)^\kappa\right)$, för $0 \le u < \infty$ och 0 för övrigt.

b) Om $\kappa = 1$ får vi $f_U(u) = \frac{1}{\lambda}\left(\frac{u}{\lambda}\right)^{1-1}\exp\left(-\left(\frac{u}{\lambda}\right)^1\right) = \frac{1}{\lambda}\exp\left(-u/\lambda\right)$, vilket vi känner igen som täthetsfunktionen för en exponentialfördelning med parametern λ.