Stockholms Universitet
Statistiska institutionen
Per Gösta Andersson

Statistikens grunder 1

SKRIFTLIG TENTAMEN
Måndagen den 30 september, 2019

Tillåtna hjälpmedel: miniräknare

Gräns för godkänt: 50 poäng av totalt 100.

För maximal poäng krävs på varje uppgift tydliga, utförliga och välmotiverade lösningar.

1. (18p) Vid tillverkningen av en viss produkt kan två typer av fel förekomma. Låt oss kalla dessa A och B. Genom långvariga mätningar har man övertygat sig om att A förekommer i 4%, att B förekommer i 8% och att både A och B förekommer i 0.5% av produkterna.

   (a) I hur stor andel av produktionen förekommer inget av felen, dvs vad är sannolikheten att en slumpmässigt vald produkt inte har något av dessa fel?

   (b) Förekommer felen oberoende av varandra?

   (c) Antag att du håller i en produkt som du vet har ett fel av typ B. Vad är då sannolikheten att den även har felet A?

2. (15p) Till en tentamensskrivning (dock ej denna!) väljs tre teoriuppgifter slumpmässigt från en lista med 32, bland vilka finns bevisen för "De stora talens lag" och centrala gränsvärdessatsen. Använd klassiska sannolikhetsdefinitionen för att beräkna nedanstående sannolikheter.

   (a) Vad är sannolikheten att inget av dessa två bevis kommer med på skrivningen?

   (b) Vad är sannolikheten att exakt ett av dessa två bevis kommer med på skrivningen?

3. (18p) En viss gen förekommer i en mycket stor population med relativa frekvensen γ, dvs om man väljer en individ slumpmässigt ur populationen, så är sannolikheten γ att individens DNA innehåller denna gen. Låt X vara antalet individer med genen i ett urval av storleken n, där individerna väljs slumpmässigt utan återläggning.

   (a) Argumentera för att det är lämpligt att anta att X är binomialfördelad. Ange också parametrarna för X.

   (b) Beräkna $P(X \geq 1)$, om $n = 10$ och $\gamma = 0.25$.

   (c) Beräkna approximativt $P(X \geq 480)$, om $n = 10\,000$ och $\gamma = 0.05$.

4. (18p) Kostnaden för att delta i ett visst lotteri är 10 kr. Vinsten i lotteriet, där kostnaden dragits bort, är en stokastisk variabel X med nedanstående sannolikhetsfördelning:

x	990	90	0	-10
$P(X = x)$	0.001	0.009	0.14	0.85

   (a) Vad är sannolikheten att vinsten blir positiv (> 0)?

   (b) Bestäm förväntad vinst ($E(X)$) och standardavvikelse ($\sqrt{V(X)}$) för vinsten.

   (c) En person spelar på lotteriet en gång i veckan under tre år ($3 \cdot 52 = 156$ veckor). Beräkna approximativt sannolikheten att den sammanlagda vinsten efter tre års spelande är positiv.

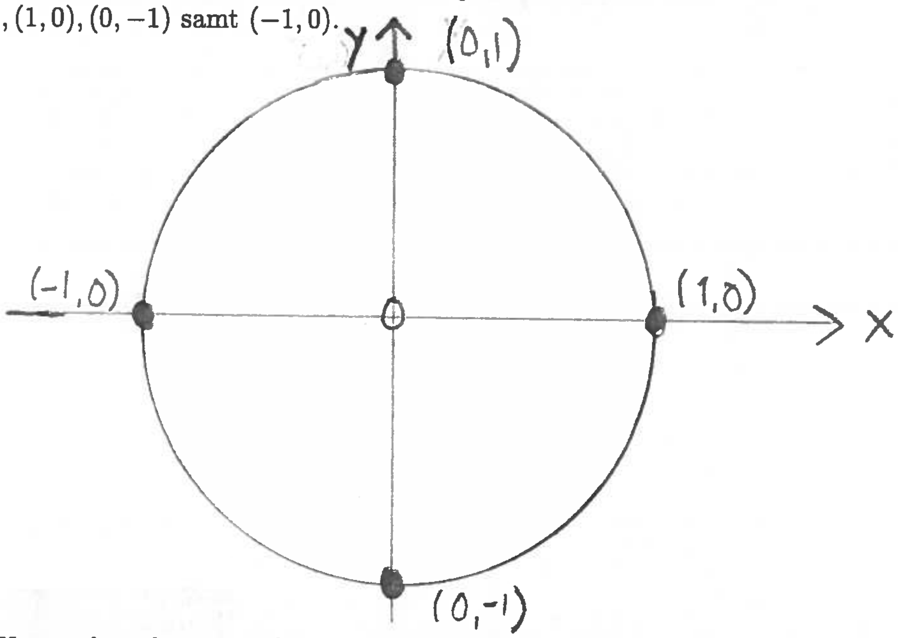
   Ledning: Beräkna väntevärde och varians för summan av vinsterna under tre års spelande och utnyttja sedan att denna summa är approximativt normalfördelad.

5. (15p) En politiker har ett auditorium som innan talet börjar består av 60% anhängare, 20% neutrala och 20% bittra motståndare. Efter talet anser 95% av anhängarna, 20% av de neutrala, men ingen av motståndarna att talet var bra.

   (a) Antag att vi väljer en person slumpmässigt bland åskådarna. Vad är sannolikheten att den personen tycker att talet var bra?

   (b) Givet att en slumpmässigt vald person tycker talet var bra, vad är sannolikheten att denna person (åtminstone före talet!) var en av de neutrala?

6. (16p) På den så kallade enhetscirkeln (se bild nedan) ska vi välja en av fyra punkter slumpmässigt. Dessa punkter har koordinaterna $(0,1),(1,0),(0,-1)$ samt $(-1,0)$.



Låt X = x-koordinaten och Y = y-koordinaten. Vi har alltså att $f_{X,Y}(0,1) = f_{X,Y}(1,0) = f_{X,Y}(0,-1) = f_{X,Y}(-1,0) = 1/4$.

(a) Beräkna kovariansen mellan X och Y, dvs $Cov(X,Y)$.

(b) Är X och Y oberoende?

(c) Kommentera kort resultaten i (a) och (b) i förhållande till varandra.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 30/9-2019

Sal: Ugglevikssalen

Tenta: Statistikens grunder 1

Kurs: Statistikens grunder

ANONYMKOD: 0080-ONU

[X] Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X	X				3
Lär.ant. 18	15	18	11	15	16				

AW

POÄNG	BETYG	Lärarens sign.
93	A	PGA

SU, STATISTIK

Skrivsal: Uggleviksssalen Anonymkod: 0080-ONU Blad nr: 1

1

$A =$ " Typ A av fel förekommer " $\rightsquigarrow P(A) = 0.04$. $\bar{A} =$ " inte fel av typ A förekommer "

$B =$ " Typ B av fel förekommer " $\rightsquigarrow P(B) = 0.08$ $\bar{B} =$ " inte fel av typ B förekommer "

$A \cap B =$ " Både typ A och B förekommer " $\rightsquigarrow P(A \cap B) = 0.005 = 0.5\%$

a) önskat: P(inget av felen) = P(inte typ A förekommer och inte typ B förekommer)

$= P(\bar{A} \cap \bar{B})$, $\bar{A} \cap \bar{B} = \overline{A \cup B}$ $\Rightarrow P(\text{inget fel}) = P(\overline{A \cup B})$

$A \cup \bar{A} = \Omega \Rightarrow P(A \cup \bar{A}) = 1 \Rightarrow P(A) = 1 - P(\bar{A})$ $\Rightarrow P(\text{inget fel}) = 1 - P(A \cup B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.04 + 0.08 - 0.005 = 0.115$

$\Rightarrow P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - 0.115 = \boxed{0.885}$

Det betyder att 88.5% av slumpmässigt valda produkter inte har något fel.

b) Om två händelser A och B är oberoende av varandra, då gäller att:

$$P(A \cap B) = P(A)\,P(B)$$

I här fallet: $P(A) \cdot P(B) = (0.04)(0.08) = 0.0032 \neq P(A \cap B) = 0.005$

Alltså A och B är beroende.

c) önskat: $P(A|B)$

Vi vet att fel typ B har förekommit, nu vill vi räkna hur stor är sannolikheten att fel typ A förekommer också.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.005}{0.08} = \boxed{0.0625}$$

6.25% av produkten med fel av typ B, förekommer med fel typ A också.

1/8

2

Vi har 3 plats som ska fyllas med 32 utfall, med utan återlämning och utan hänsyn till ordning:

Alla möjliga antal att välja 3 av 32 är: $\binom{32}{3} = \frac{32!}{29!\,3!} = \boxed{4960}$

a) Vi tar bort två bevisen från utfallsrummet, då vi har 30 utfall nu.

Bland detta 30 utfall väljer vi 3 utan återlämning och hänsyn till ordning.

Då antal möjliga utfall är: $\binom{30}{3} = \frac{30!}{27!\,3!} = 4060$

Enligt klassiska sannolikhets definition:

$$\text{Sannolikhet att händelse A inträffar} = \frac{\text{Antal av möjliga utfall som gör händelse A inträffar}}{\text{Alla möjliga utfall.}}$$

Om A: "Inget av två bevisen kommer"

$$\Rightarrow P(A) = \frac{\binom{30}{3}}{\binom{32}{3}} = \frac{4060}{4960} = \boxed{0.818}$$

b) B: "Exakt ett av dessa två bevis kommer"

B: "'Stora talen' kommer och inte 'CGV' eller 'CGV' kommer och inte 'stora talen'"

Vi har 3 platser, i första vi har bara två möjlig utfall, 'stora talen' eller 'CGV', vi väljer en av 2: $\binom{2}{1}$, vi fyller andra och tredje plats med 30 utfall: $\binom{30}{2}$

Då antal möjliga utfall för $B = \binom{2}{1}\binom{30}{2} = 870$

$$\Rightarrow P(B) = \frac{\text{antal möjliga utfall för B}}{\text{alla möjliga utfall}} = \frac{870}{4960} = \boxed{0.175}$$

1/5

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: 0080-ONU Blad nr: 2

3 a) Vi kan definera en binär variabel X_i: "en individ har viss gen."

Då vi kan säga att utfallsrummet för X_i är $\Omega_{X_i} = \{0, 1\}$ som 0 betyder individen inte har viss genen och 1 betyder individen har viss genen.

Om vi upprepar försöket att tar slumpmässiga individer från ett urval av storleken "n", då X: "antal individer som har viss gen." blir $\sum_{i=1}^{n} X_i$

som $\Omega_X = \{0, 1, 2, \dots, n\}$. $0 \in \Omega_X$ betyder att ingen av individer i urval med storlek "n" har genen. $n \in \Omega_X$ betyder att alla i urvalet har genen, osv.

Vi kan definera en binomialfördelning för X med parametrerna n, γ.

"n" är storleken av urvallet och det största värde som X antar, och "γ" är den sannolikheten som varje individ har för att ha genen.

Vi kan beteckna att:

$$X \sim bin(n, \gamma) \qquad f_X(x) = \binom{n}{x} (\gamma)^x (1-\gamma)^{n-x}$$

b)

$$X \sim bin(10, 0.25) \rightarrow f_X(x) = \binom{10}{x} (0.25)^x (0.75)^{10-x}$$

$$\left.\begin{array}{l} P(X \geq 1) = 1 - P(X = 0) \\ P(X=0) = f_X(0) = \binom{10}{0}(0.25)^0 (0.75)^{10} = 0.05631 \end{array}\right\} \Rightarrow P(X \geq 1) = 1 - 0.05631 = \boxed{0.94369}$$

c) $X \sim bin(10000, 0.05)$

$$\left.\begin{array}{l} np = 500 > 5 \\ n(1-p) = 9500 > 5 \end{array}\right\} \Rightarrow \text{vi kan approximera fördelningen till } N(np, np(1-p))$$

$$X \sim bin(10000, 0.05) \overset{appr}{\sim} N(500, 475) \rightarrow Z = \frac{X - 500}{\sqrt{475}} \sim N(0,1)$$

$$P(X \geq 480) = 1 - P(X \leq 480) = 1 - P\left(\frac{X - 500}{\sqrt{475}} \leq \frac{480 - 500}{\sqrt{475}}\right) = 1 - P(Z \leq -0.92)$$

$$= 1 - \Phi(-0.92) = \Phi(0.92) = 0.82121$$

$$\Rightarrow P(X \geq 480) = \boxed{0.82121}$$

1/8

4)

x	990	90	0	-10
$P(X=x_i)$	0.001	0.009	0.14	0.85

a) $P(X>0) = P(X=90) + P(X=990) = (0.009) + (0.001) = \boxed{0.01}$

b) $E(X) = \sum_{i=1}^{4} x_i f(x_i) = (990)(0.001) + (90)(0.009) + (0)(0.14) + (-10)(0.85) = \boxed{-6.7}$

$V(X) = \sum_{i=1}^{4} x_i^2 f(x_i) - (E(X))^2 = (990)^2(0.001) + (90)^2(0.009) + (-10)^2(0.85) - (-6.7)^2 = \boxed{1093.11}$

Standardavvikelse $= \sqrt{V(X)} = \sqrt{1093.11} \approx \boxed{33.062}$

c) $P(X>0) = P(\text{vinsten är positiv}) \overset{4a}{=} 0.01$

Det kan defineras en bionomialfördelning för Y: "Får positiv vinst efter 3 år (156 veckor)"

Personen spelar 156 lotteri med sannolikhet av 0.01 för positiv vinst för varje lotteri.

Men vi söker en sannolikhet kopplad till den sammanlagda vinsten.

$Y \sim bin(156, 0.01)$
$np = (156)(0.01) = 1.56$
$n(1-p) = 154.44 > 5$
$\Rightarrow$ approximation $Y \sim N(np, npq)$

$\Rightarrow Y \sim bin(156, 0.01) \xrightarrow{\text{approx.}} Y \sim N(1.56, 1.5444)$

$P(Y>0) = P\left(\frac{Y - 1.56}{\sqrt{1.5444}} > \frac{0 - 1.56}{\sqrt{1.5444}}\right) = P(Z \geq -1.26) = \Phi(1.26) = \boxed{0.89617}$

$\Rightarrow P(Y>0) = \boxed{0.89617}$ ✓ $\rightarrow$ sannolikhet att sammanlagda vinsten blir positiv efter 3 år (156 veckor).

rimligt resultat?

/ 11

SU, STATISTIK

Skrivsal: Uggleviksssalen **Anonymkod:** 0080-ONU **Blad nr:** 3

5

A: "anhängare" $\rightsquigarrow$ $P(A) = 0.6 = 60\%$

N: "Neutrala" $\rightsquigarrow$ $P(N) = 0.2 = 20\%$

M: "Motståndare" $\rightsquigarrow$ $P(M) = 0.2 = 20\%$

B: "Talet är bra"

$P(B|A) = 0.95$ $P(B|N) = 0.20$; $P(B|M) = 0$

a) $P(B) = ?$

Enligt totala sannolikhetssatsen: $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$

$P(\sum_{i=1}^{n} A_i) = 1$, och $A_1, \dots, A_n$ är disjunkta.

Här har vi $P(A) + P(N) + P(M) = 1$

$$P(B) = P(B|A)P(A) + P(B|N)P(N) + P(B|M)P(M)$$

$$\Rightarrow P(B) = (0.95)(0.6) + (0.2)(0.2) + (0)(0.2) = \boxed{0.61}$$

61% av personer som deltog auditorium tycker att talet var bra.

b) önskat: $P(N|B) = ?$

$$P(N|B) = \frac{P(N \cap B)}{P(B)} = \frac{P(B|N)P(N)}{P(B)} = \frac{(0.2)(0.2)}{0.61} \approx \boxed{0.066}$$

Det betyder att 6.6% av de som tycker att talet var bra är neutrala.

/15

6

Y \ X	-1	0	1	
-1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

a)

$E(X) = \sum x_i f_X(x_i) = (-\frac{1}{4}) + 0 + (\frac{1}{4}) = \boxed{0}$

$V(X) = \sum x_i^2 f_X(x_i) - E^2(X) = \frac{1}{4} + 0 + \frac{1}{4} = \boxed{\frac{1}{2}}$

$E(Y) = \sum y_i f_Y(y_i) = \frac{-1}{4} + 0 + \frac{1}{4} = \boxed{0}$

$V(Y) = \sum y_i^2 f_Y(y_i) = \frac{1}{4} + 0 + \frac{1}{4} = \boxed{\frac{1}{2}}$

$Cov(X,Y) = E(XY) - E(X)E(Y) = E(XY) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} xy f_{X,Y}(x,y)$

$Cov(X,Y) = E(XY) = (-1)(-1)(0) + (-1)(1)(0) + (-1)(1)(0) + (1)(1)(0) = 0$

b) För att X och Y vara oberoende, då måste gällas

$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ för alla $x \in \Omega_X$ och $y \in \Omega_Y$

$f_X(-1) \cdot f_Y(-1) = (\frac{1}{4})(\frac{1}{4}) = \frac{1}{16} \neq f_{X,Y}(-1,-1) = 0 \Rightarrow$ X och Y är beroende.

c) Om två händelsen är oberoende då kovariansen mellan händelserna är lika med 0, tex A och B oberoende $\Rightarrow Cov(A,B) = 0$

Men om $Cov(A,B) = 0$, inte betyder att A och B är oberoende.

Som vi ser resultaten i 6a att $Cov(X,Y) = 0$, men de är beroende enligt 6b. Vi kan säga att trots att X och Y är beroende, kovariansen mellan dem är 0. $Cov(X,Y) = 0$ betyder att $\rho_{X,Y} = 0 \rightarrow$ det finns ingen samband mellan X och Y, men de kan vara beroende av varandra.

/6