

Extrauppgifter i matematisk statistik

BT 2014

1. Mängden A är dubbelt så sannolik som B . Hur förhåller sig $P(A|B)$ till $P(B|A)$?
2. Två händelser A och B har sannolikheter skilda från noll.
 - (a) A och B är disjunkta. Kan A och B vara oberoende?
 - (b) A och B är oberoende. Kan A och B vara disjunkta?
3. Sannolikheten för att en satellit träffas av en meteorit under ett varv runt jorden är 0.025. Låt X vara antalet träffar efter 100 varv.
 - (a) Vad har X för fördelning?
 - (b) Beräkna $P(X \geq 4)$
 - (c) Låt Y vara antal varv tills satelliten träffas för första gången. Vad har Y för fördelning? Beräkna $\text{Exp}[Y]$.
4. Låt A och B vara två händelser. Visa att (t.ex. med hjälp av Venn-diagram)
 - (a) $B = (B \cap A) \cup (B \cap A')$
 - (b) $A \cup B = A \cup (B \cap A')$
5. I genomet för *Mus musculus* (husmus) finns det ungefär 24 000 gener och av dessa har 3 000 ingen känd funktion. Antag att vi väljer 1 000 av de 24 000 generna på måfå och låter X =antalet valda gener utan känd funktion. Vad har X för fördelning?
6. Antag att $X \sim \text{Geo}(1/2)$. Rita fördelningsfunktionen F_X .
7. Låt X vara en stokastisk variabel med täthetsfunktion $f_X(x) = 3x^2, 0 \leq x \leq 1$. Beräkna fördelningsfunktionen F_X .
8. Låt X vara en stokastisk variabel och a och b vara tal där $a > 0$.

- (a) Använd definitionen för väntevärde och visa att $\text{Exp}[aX + b] = a \text{Exp}[X] + b$.
 - (b) Använd definitionen för varians samt (a) för att visa att $\text{Var}[aX + b] = a^2 \text{Var}[X]$.
9. Visa att $\text{Var}[X] = \text{Exp}[X^2] - (\text{Exp}[X])^2$.
10. Låt X och Y vara två kontinuerliga oberoende stokastiska variabler med täthetsfunktioner $f_X(x)$ och $f_Y(y)$. Visa att $\text{Exp}[XY] = \text{Exp}[X] \text{Exp}[Y]$.
11. En tillverkare av bildäck tänker lansera en ny typ av radialdäck. Genom mätningar har man kommit fram till att livslängder (i miles) beskrivs av normalfördelade stokastiska variabler med parametrarna $\mu = 36500$ och $\sigma^2 = 5000$. Tillverkar vill kunna garantera en livlängd L i miles. Hur ska man välja L om man vill att 90% av däcken ska överleva garantinivån?
12. Antag att den stokastiska variabeln X har följande täthetsfunktion

$$f_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0$$

med parametrarna $\alpha = 1$ och $\beta = 2$. Beräkna $P(1 \leq X \leq 3)$. För α heltal gäller att $\Gamma(\alpha) = (\alpha - 1)!$.

13. Visa att sannolikheten för att exakt en av händelserna A och B inträffar är $P(A) + P(B) - 2P(A \cap B)$.
14. Den stokastiska variabeln X kan bara anta värdena 3, 4, 7, 8 och 9. Man känner följande värden på sannolikhetsfunktionen:

$$p_X(3) = 1/3, p_X(4) = 1/4, p_X(7) = 1/6, p_X(8) = 1/6.$$

Beräkna

- (a) $p_X(9)$,
- (b) $F_X(5)$,
- (c) $P(4 \leq X \leq 8)$ och $P(X \geq 8)$.

15. Låt den tvådimensionella diskreta stokastiska variabeln (X, Y) ha sannolikhetsfunktionen

$$p_{X,Y}(x, y) = \frac{x^2 + y^2}{32}, \text{ för } x = 0, 1, 2, 3 \text{ och } y = 0, 1.$$

Beräkna $\text{Exp}[X]$.

16. Linus och Linnéa har i en världsberömd lustträdgård plockat sju frukter av aptitligt utseende. De är lyckligt ovetande om att tre av frukterna är giftiga. Om Linus sätter i sig fyra frukter på måfå och Linnéa de återstående tre, hur stor är sannolikheten för att båda blir förgiftade?

17. (X, Y) har täthetsfunktionen

$$f_{XY}(x, y) = \begin{cases} e^{-y} & \text{om } 0 < x < y, \\ 0 & \text{annars.} \end{cases}$$

- (a) Bestäm de marginella täthetsfunktionerna för X och Y .
 - (b) Beräkna $P(X > 2 | < 4)$.
 - (c) Är X och Y oberoende?
18. Tennisspelarna Linus och Linnéa skall spela en match mot varandra. Händelsen att Linnéa vinner ett set är oberoende av utgången av tidigare set och sannolikheten för detta är 0.6. Matchen slutar när en av spelarna har vunnit tre set. Låt X vara det totala antalet set i matchen. Beräkna
- (a) $p_X(x)$,
 - (b) $\text{Exp}[X]$.
19. Den stokastiska variabeln $Y \sim t_{10}$. Bestäm c så att $P(-c < Y < c) = 0.95$.
20. Vid tillverkning av en viss sorts färg tillsätts färgpigmentet med hjälp av en doseringsapparat, men man är inte säker på om den har tillräcklig precision. Om man ställer in apparaten på dosen μ blir den verkliga dosen $X = \mu + \epsilon$, där $\epsilon \sim N(0, \sigma^2)$. Man vill att minst 99% av burkarna ska ha en acceptabel pigmentdos, dvs en dos i intervallet $(\mu - 0.2, \mu + 0.2)$, i det långa loppet.

Man genomförde 51 provdoseringar med $\mu = 10$ och fick stickprovsstandardavvikelsen $s = 0.0629$. Uppfyller apparaten precisionskravet vid $\mu = 10$? Besvara frågan med hjälp av ett lämpligt 95 % konfidensintervall.

21. För att jämföra effekterna hos tre olika blodtryckssänkande mediciner behandlades tre grupper om vardera 10 patienter med de olika medicin-
eran. Efter en månad mättes sänkningen av blodtrycket.

	Medelvärde	Stickprovsstandard- avvikelse
Medicin 1	17.3	6.19
Medicin 2	21.1	7.26
Medicin 3	10.8	5.23

Modell: Tre oberoende stickprov från $N(\mu_i, \sigma^2)$, $i = 1, 2, 3$.

Konstruera ett konfidensintervall för σ av typen $(0, a)$ med konfidensgrad 0.95.

22. Betjäningstiderna för ett kösystem är exponentialfördelade med väntevärde μ . Man har observerat 80 oberoende betjäningstider och fått medelvärdet $\bar{x} = 4.5$ minuter.
- (a) Konstruera ett konfidensintervall för μ med konfidensgraden approximativt 0.95.
 - (b) Låt p vara sannolikheten att en betjäningstid är mer än tio minuter. Konstruera ett konfidensintervall för p med konfidensgraden approximativt 0.95.
23. På misstänkta rattfyllerister gör man tre bestämningar av alkoholhalten i blodet. Resultaten x_1 , x_2 och x_3 antas utgöra ett slumpmässigt stickprov från $N(\mu, 0.05^2)$ där μ är den verkliga alkoholhalten i blodet. Om $\mu > 0.2$ har personen gjort sig skyldig till rattonykterhet. Låt oss anta att den domstol som skall döma tar hänsyn till osäkerheten i

mätningarna genom att beräkna det aritmetiska medelvärdet $\bar{x}$ av de tre analysresultaten och därefter förklara personen skyldig om

$$\bar{x} > 0.2 + \lambda_0 \cdot 0.01 \cdot 0.05 / \sqrt{3}$$

men oskyldig annars. Med statistisk terminologi kan man säga att domstolen prövar $H_0 : \mu = 0.2$ mot $H_1 : \mu > 0.2$ på nivån 0.01. Vilka av följande påståenden ger en någorlunda korrekt beskrivning av vad som kommer hända i det långa loppet?

- (a) högst 1% av alla frikända är skyldiga
- (b) högst 1% av alla oskyldiga blir dömda
- (c) högst 1% av alla skyldiga blir frikända
- (d) högst 1% av alla dömda är oskyldiga

24. Med en viss mätutrustning registreras den radioaktiva bakgrundsstrålningen på en ort. Det är rimligt att anta att antalet registrerade partiklar under t minuter är $Po(\lambda t)$ där $\lambda = 5$ (min^{-1}). Efter ett radioaktivt utsläpp misstänker man att strålningen har ökat. Hur länge behöver man mäta strålningen om man vill pröva

$$H_0 : \lambda = 5 \quad \text{mot} \quad H_1 : \lambda > 5$$

på nivån 0.01 med ett test som ger utslag med säkerheten 0.99 om strålningen ökat 50%. Lämplig approximation får utnyttjas.

25. Man har mätt en viss föroreningshalt i jordprover tagna dels i närheten av en industri och dels i ett "rent område" med motsvarande jordmån. Det förutsätts att en mätning ger ett lognormalfördelat värde och därför är nedanastående medelvärden beräknade för de logaritmerade mätvärdena.

	antal obs.	medelvärde	standardavvikelse
Nära industrin	8	1.8	0.49
Rent område	9	1.1	0.46

Undersök om föroreningshalten kan anses vara högre i det förorenade området.

26. Ett företag tillverkar reläer vid fyra olika fabriker; A, B, C och D. För att undersöka om fabrikerna har samma kvalitet på sina reläer undersökte man 200 enheter från var och en av fabrikerna och räknade antalet defekta och korrekta enheter.

	A	B	C	D
Defekta	6	4	4	16
Korrekta	194	196	196	184

Det förefaller rimligt att anta att A, B och C har samma defektsannolikhet p_1 medan D har en annan defektsannolikhet p_2 . Konstruera ett 95 % konfidensintervall för $p_1 - p_2$.

27. Genotyperna AA , Aa och aa för en viss allel förekommer med sannolikhet θ^2 , $2\theta(1 - \theta)$ och $(1 - \theta)^2$ i en population. Man har på måfå valt ut ett antal individer och fått följande resultat:

Typ	AA	Aa	aa
Antal	34	50	19

ML-skatta andelen A -gener i populationen, dvs θ .

28. Använd ML-skattningen av θ från uppgift 27 och utför ett χ^2 -test för att pröva om den angivna modellen är rimlig.
29. En kaffeleverantör påstår att mängden kaffe i ett 500g-paket kan betraktas som en stokastisk variabel $X \sim N(500, 10^2)$. Vid en undersökning av 100 paket fick man följande resultat:

Vikt	< 495 g	$495 - 504$ g	> 505 g
Antal paket	39	45	16

Kaffemängder i olika paket kan anses vara oberoende stokastiska variabler.

Pröva med hjälp av ett χ^2 -test på nivån 0.01 hypotesen att kaffemängderna är normalfördelade som leverantören påstår.

30. I följande tabell är y_i observerade värden på en stokastisk variabel $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, där $\epsilon_1, \dots, \epsilon_n$ är oberoende och $N(0, \sigma^2)$.

x_i	1	2	3	4	5	6	7	8	9	10
y_i	11.5	9.8	7.0	5.5	3.9	2.0	0.4	-1.4	-3.1	-5.0

- (a) Beräkna punktskattningarna av β_0 , β_1 , och σ .
- (b) Rita in värdena tillsammans med den beräknade regressionslinjen i ett koordinatsystem.