Tentamensskrivning i Matematisk statistik för **D3**, TMA290
Onsdagen 13 mars 2002 kl 08.45 – 12.45 i sal V
Lärare: Dan Mattsson, tfn 772 5349
Jour: Kristina Wärmefjord, tfn 772 4996
Hjälpmedel: Utdelad formelsamling med tabeller (även BETA, Physics Handbook, skoltabeller, till exempel TEFYMA). Valfri räknedosa utan kommunikationsmöjlighet med andra räknedosor samt med *tömda* minnen. Inga egna anteckningar eller lärobok.
Betygsgränser: Maxpoäng 30, för betyg 3, 4 och 5 erfordras 12, 18 respektive 24 poäng.

---

1. Funktioner av stokastiska variabler är i sin tur nya stokastiska variabler. Låt X vara en stokastisk variabel med strängt växande fördelningsfunktion F_X. Sätt $Y = F_X(X)$. Vad har Y för fördelning? (3p)

2. Härled konfidensintervall med konfidensgraden $1 - \alpha$ för väntevärdet i normalfördelningen $\mathrm{N}(\mu, \sigma^2)$, i fallet då σ^2 är okänd. (3p)

3. Vid sändning av datapaket med ett visst protokoll upptäcks det, och korrigeras för, paket med färre än tre felaktiga bitar. Vid ett tillfälle vill Arne skicka 5 paket över nätverket. Hur stor sannolikhet är det att alla paketen kommer fram korrekta om varje paket är 1024 bitar? Man får anta att varje bit sänds oberoende av de övriga och att sannolikheten för fel på en bit är $p = 0.0005$. (3p)

4. Vi har tre skrivare och varje utskrift dirigeras till skrivare ett, två eller tre med sannolikhet 0.6, 0.3 och 0.1, respektive, oberoende av varandra. Sannolikheten för papperstrassel är för varje utskrift 0.01, 0.05 och 0.04 för de olika skrivarna.

   Om en utskrift får papperstrassel vad är sannolikheten för att det var den första skrivaren som strulade? (3p)

5. Av 193 oberoende intervjuer ställde sig 75 positiva till ett förslag om rökförbud på krogen. Ställ upp ett 95% konfidensintervall för den sanna andelen positiva i populationen. (3p)

6. Två tillverkningsprocesser kan användas för att härda kopparrör, den enda väsentliga skillnaden är den nödvändiga temperaturen vid processen. För att jämföra metoderna delas 8 stycken kopparrör med olika diametrar i två bitar, där en slumpmässigt vald bit av ett rör gräddas i högre temperatur, den andra i lägre. För dessa 8 par av rörbitar mätte man sedan upp draghållfastheten:

Par	Låg temp.	Hög temp.
1	16.5	16.8
2	17.6	17.7
3	16.9	17.1
4	15.8	15.9
5	18.4	18.6
6	17.5	18.0
7	17.6	17.7
8	16.1	16.0

   Använd dessa observationer för att testa hypotesen att högtemperatursmetoden ger högre draghållfasthet. (3p)

Vänd!

7. För att avgöra hur bilar anländer till en parkeringsplats studeras antalet ankommande bilar i 100 stycken femminuters intervall. Följande tabell erhölls

Antal bilar	0	1	2	3	4	≥ 5
Frekvens, o_i	7	30	30	18	10	5

och medelvärdet på antalet anländande bilar i ett femminuters intervall var 2.12. Finns det belägg för en hypotes om att bilarna anländer som en Poissonprocess? (3p)

8. Låt X_1 vara antalet motorcyklar med en person på, som passerar en punkt på motorvägen under en tidsperiod t. Antag att dessa motorcyklar kommer enligt en Poissonprocess med intensitet $\lambda_1 = 0.85$ MC/minut. Motorcyklar med två personer på, passerar samma punkt enligt en Poissonprocess med intensitet $\lambda_2 = 0.20$ MC/minut oberoende av den första processen. Bestäm sannolikheten att det under 2 minuter passerar exakt tre motorcykelburna *personer*.

   Givet att det passerar exakt n *motorcyklar* bestäm fördelningen för antalet av dessa som hade en person på sig. Dvs, ställ upp ett uttryck för $\mathsf{P}(X_1 = k | X_1 + X_2 = n)$ för de möjliga värdena på k. (3p)

9. Paretofördelningen används för att modellera bland annat sprickbildning, där sannolikhetstätheterna har långsamt avtagande svansar, dvs stora värden har ganska stora sannolikheter. Täthetsfunktionen ges av

$$f_X(x) = \theta\, c^\theta\, x^{-\theta-1} \quad x \geq c,\ \theta > 1.$$

   Antag att $c > 0$ är given och bestäm *momentmetodens* skattning av θ för stickprovet $X_1, \ldots, X_n$. Beräkna sedan skattningens värde för observationerna

$$2.5821 \quad 2.5760 \quad 3.6077 \quad 2.8172 \quad 2.7236 \quad 4.6739,$$

   om $c = 2$. (3p)

10. En modell för att modellera aktiekurser är att om aktiekursen vid tiden t är den stokastiska variabeln X_t, så antas $\log X_t \sim \mathrm{N}(\mu + c \cdot t, \sigma^2 t)$. Kursen för en ELUX A-aktie idag $X_0 = 190$ är känd. Antag att $c = 0.01$ och $\sigma = 0.1$ (dessa beror naturligtvis på aktien och är svåra att skatta) och beräkna väntevärdet för kursen om ett år, dvs $\mathsf{E}[X_1]$. (3p)

1. Om F_X är en strikt växande funktion så existerar dess invers F_X^{-1} (också strikt växande), så att $F_X^{-1}(F_X(x)) = x$. Då får vi

$$F_Y(t) = \mathsf{P}(Y \leq t) = \mathsf{P}(F_X(X) \leq t) = \mathsf{P}\left(X \leq F_X^{-1}(t)\right) = F_X(F_X^{-1}(t)) = t,$$

dvs $f_Y(t) = 1$ då $0 \leq t \leq 1$. Y har likformig fördelning på intervallet $[0, 1]$.

2. Om $X_1, \ldots, X_n$ är ett stickprov på en normalfördelad stokastisk variabel med väntevärde μ och varians σ^2, så är $\overline{X}$ normalfördelad med väntevärde μ och varians σ^2/n. Alltså

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathrm{t}_{n-1}.$$

Vi kan ur t_{n-1} tabeller bestämma $t_{\alpha/2}$ så att

$$\mathsf{P}\left(-t_{\alpha/2} \leq T \leq t_{\alpha/2}\right) = 1 - \alpha.$$

Så med sannolikhet $1 - \alpha$ får vi att

$$-t_{\alpha/2} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2}$$

eller

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}.$$

3. Låt X vara antalet felaktiga bitar i ett paket, $X \sim \mathrm{Bin}(n, p)$. Då är sannolikheten att ett paket kommer fram korrekt

$$q = \mathsf{P}(X < 3) = \sum_{k=0}^{2} \binom{1024}{k} p^k (1-p)^{1024-k} = 0.9847$$

Av $m = 5$ översända paket, är antalet korrekta paket, $Y \sim \mathrm{Bin}(m, q)$, och

$$\mathsf{P}(Y = m) = q^m = 0.926.$$

4. Låt A, B och C beteckna händelserna att skrivare A, B och C handhar utskriften. Låt J beteckna händelsen att vi får papperstrassel. Givet är att $\mathsf{P}(A) = 0.6$, $\mathsf{P}(B) = 0.3$ och $\mathsf{P}(C) = 0.1$, samt att trasselsannolikheterna är $\mathsf{P}(J|A) = 0.01$, $\mathsf{P}(J|B) = 0.05$ och $\mathsf{P}(J|C) = 0.04$.

   Då är $\mathsf{P}(J) = \mathsf{P}(J|A)\,\mathsf{P}(A) + \mathsf{P}(J|B)\,\mathsf{P}(B) + \mathsf{P}(J|C)\,\mathsf{P}(C) = 0.025$ och $\mathsf{P}(A|J) = \mathsf{P}(A \cap J)/\mathsf{P}(J) = \mathsf{P}(J|A)\,\mathsf{P}(A)/\mathsf{P}(J) = 0.24$.

   På samma sätt kan man räkna ut att $\mathsf{P}(B|J) = 0.60$ och $\mathsf{P}(C|J) = 0.16$.

5. Låt $X \sim \mathrm{Bin}(n, p)$. Vi skattar p med $\widehat{p} = X/n$ som uppfyller $\mathsf{E}[\widehat{p}] = p$ och $\mathrm{Var}(\widehat{p}) = p(1-p)/n$. Centrala gränsvärdessatsen (och Slutsky) ger att

$$Z = \frac{\widehat{p} - p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \overset{\text{approx}}{\sim} \mathrm{N}(0, 1)$$

så om vi bestämmer $z_{\alpha/2}$ så att $\mathsf{P}\left(|Z| \leq z_{\alpha/2}\right) = 1 - \alpha$, gäller med sannolikhet $1-\alpha$ att

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p \leq \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

Observationer $n = 193$ och $x = 75$ ger punktskattningen $\widehat{p} = x/n = 0.3886$ av p. Med $1 - \alpha = 0.95$ får vi $z_{\alpha/2} = 1.96$ så ett observerat konfidensintervall för p blir

$$0.3535 \leq p \leq 0.4237 \quad (95\%).$$

6. Låt X_i vara draghållfastheten hos rör i efter gräddning i den lägre temperaturen och Y_i motsvarande för den högre. Vi ansätter modellen att skillnaden i draghållfasthet mellan de olika gräddningsförfarandena

$$V_i = Y_i - X_i$$

är oberoende och $\mathsf{E}[V_i] = \mu$ och $\mathrm{Var}(V_i) = \sigma^2$. Vi vill testa

$$H_0 : \mu \leq 0 \quad \text{mot} \quad H_1 : \mu > 0$$

på signifikansnivå $\alpha = 0.05$. Vi har följande observerade skillnader

$$0.3 \quad 0.1 \quad 0.2 \quad 0.1 \quad 0.2 \quad 0.5 \quad 0.1 \quad -0.1$$

med observerat medelvärde $\overline{v} = 0.175$ och standardavvikelse $s_v = 0.1753$. Med hjälp av centrala gränsvärdessatsen är under H_0

$$T = \frac{\overline{V}}{S/\sqrt{n}} \overset{\text{approx}}{\sim} t_{n-1}\text{-fördelat}$$

och vi förkastar H_0 till förmån för H_1 för stora värden på T. Ur t_7-tabell bestämmer vi $t_\alpha = 1.8946$ så att

$$\mathsf{P}\left(T > t_\alpha | H_0\right) = 0.95.$$

Vi observerar utfallet på T till $t = 2.8243 > t_\alpha$ och förkastar H_0 på nivå $\alpha = 5\%$. (p-värde 1.28%)

7. Låt X vara antalet bilar under ett femminutesintervall. Vi skall testa

$$H_0 : X \sim \mathrm{Po}(c) \quad \text{mot} \quad H_1 : \text{inte } H_0.$$

Med medelantalet bilar skattar vi c i Poissonfördelningen under H_0, $\widehat{c} = \overline{X}$, med observerat värde $\widehat{c} = 2.12$. Vi räknar ut skattade sannolikheter $\widehat{p}_i$ att ett intervall innehåller i bilar.

$$\widehat{p}_i = \widehat{\mathsf{P}\left(X = i\right)} = \frac{\widehat{c}^i}{i!}\mathrm{e}^{-\widehat{c}}$$

och på $n = 100$ observerade intervall förvänta oss $e_i = np_i$ som skattas med $\widehat{E}_i = n\widehat{p}_i$, med observerade värden

Antal bilar	0	1	2	3	4	≥ 5
$\widehat{e}_i = n\widehat{p}_i$	12	25.4	26.9	19	10.1	6.4
Frekvens, o_i	7	30	30	18	10	5

Vi bildar sedan teststatistikan

$$Q = \sum_{i=1}^{6} \frac{(O_i - \widehat{E}_i)^2}{\widehat{E}_i} \qquad q = \sum_{i=1}^{6} \frac{(o_i - \widehat{e}_i)^2}{\widehat{e}_i} = 3.63$$

som under H_0 är approximativt χ^2_4-fördelat. Förkasta för stora värden på Q. Ur χ^2_4-tabell har vi att $\mathsf{P}\left(\chi^2_4 > q_\alpha\right) = 0.10$ för $q_\alpha = 7.78$. Vi kan inte förkasta H_0 på nivå $\alpha = 0.10$. (p-värde 0.46). Vi har således inte någon statistisk säkerställd skillnad från en Poissonfördelning.

8. Låt X_1 vara antalet motorcyklar med en person under de två minuterna, $X_1 \sim \text{Po}(c_1)$, $c_1 = \lambda_1 t = 1.70$, och $X_2 \sim \text{Po}(c_2)$, $c_2 = 0.40$, är motsvarande för MC med två personer. Den sökta sannolikheten är då, med utnyttjande av oberoendet,

$$\mathsf{P}(X_1 = 3,\ X_2 = 0) + \mathsf{P}(X_1 = 1,\ X_2 = 1) = \frac{c_1^3}{3!}\mathrm{e}^{-c_1}\frac{c_2^0}{0!}\mathrm{e}^{-c_2} + \frac{c_1^1}{1!}\mathrm{e}^{-c_1}\frac{c_2^1}{1!}\mathrm{e}^{-c_2} = 0.184.$$

Givet att n motorcyklar passerat söker vi $\mathsf{P}(X_1 = k | X_1 + X_2 = n)$ för $k = 0, 1, \ldots, n$. Definition av betingning och direkt uträkning ger

$$\begin{aligned} \mathsf{P}(X_1 = k | X_1 + X_2 = n) &= \frac{\mathsf{P}(X_1 = k, X_2 = n-k)}{\mathsf{P}(X_1 + X_2 = n)} = \frac{\frac{c_1^k}{k!}\mathrm{e}^{-c_1}\frac{c_2^{n-k}}{(n-k)!}\mathrm{e}^{-c_2}}{\frac{(c_1+c2)^n}{n!}\mathrm{e}^{-(c_1+c2)}} \\ &= \binom{n}{k}\left(\frac{c_1}{c_1+c_2}\right)^k\left(\frac{c_2}{c_1+c_2}\right)^{n-k}, \end{aligned}$$

det vill säga binomialfördelat med $p = c_1/(c_1 + c_2) = \lambda_1/(\lambda_1 + \lambda_2)$.

9. Vi får momentskattningen genom att lösa ekvationen $\mathsf{E}[X] = \overline{x}$. Integrering ger

$$\mathsf{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx = \int_c^{\infty} \theta c^\theta x^{-\theta}\,dx = \frac{\theta}{\theta - 1}c,$$

varur vi får $\theta = \overline{x}/(\overline{x} - c)$ eller $\widehat{\theta} = \overline{X}/(\overline{X} - c) = 2.72$.

10. Med $Y_t = \log X_t$ vill vi bestämma $\mathsf{E}[X_1] = \mathsf{E}\left[e^{Y_1}\right]$, där Y_t är normalfördelad $\sim \text{N}(\mu + c \cdot t, \sigma^2) = \text{N}(\mu_Y, \sigma^2)$. Vi söker alltså den momentgenererande funktionen $m_{Y_1}(s)$ i punkten $s = 1$. Antingen kommer man ihåg att $m_{Y_t}(s) = \mathrm{e}^{\mu_Y s + \sigma^2 s^2/2}$ och

$$\mathsf{E}[X_t] = m_{Y_t}(1) = \mathrm{e}^{\mu + ct + \sigma^2/2}.$$

Alternativt bestämmer man den momentgenererande funktionen för en stokastisk variabel $Z \sim \text{N}(0, 1)$

$$\begin{aligned} m_Z(s) &= \mathsf{E}\left[\mathrm{e}^{sZ}\right] = \int_{-\infty}^{\infty} \mathrm{e}^{sz}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,dz = \mathrm{e}^{s^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\eta}}\mathrm{e}^{-(z^2-2sz+s^2)/2}\,dz \\ &= \mathrm{e}^{s^2/2}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(z-s)^2/2}\,dz}_{=1} = \mathrm{e}^{s^2/2}, \end{aligned}$$

eftersom integranden är tätheten för en $\text{N}(s, 1)$ fördelad stokastisk variabel. Nu är

$$m_{Y_t}(s) = \mathsf{E}\left[e^{sY_t}\right] = \mathsf{E}\left[e^{s(\mu_Y + \sigma Z)}\right] = \mathrm{e}^{s\mu_Y} m_Z(s\sigma) = \mathrm{e}^{s\mu_Y}\mathrm{e}^{(s\sigma)^2/2} = \mathrm{e}^{s(\mu+ct)+\sigma^2 s^2/2}.$$

Sålunda med $s = 1$ får vi

$$\mathsf{E}[X_t] = \mathrm{e}^{\mu+ct+\sigma^2 t/2} = \mathrm{e}^{\mu}\cdot\mathrm{e}^{ct+\sigma^2 t/2} = X_0 \cdot \mathrm{e}^{ct+\sigma^2 t/2},$$

och $\mathsf{E}[X_1] = 192.87$.