Lösningar till tentamen i Statistik för Fysiker m. fl
MSN 220
Lördagen den 6 mars 2004

1. Låt S beteckna händelsen att en slumpvis vald person är sjuk och D att en person diagnosticeras som sjuk. Uppgiftens förutsättningar är att

$$\mathbb{P}(S) = 10^{-4}, \mathbb{P}(D|S) = 0,99, \mathbb{P}(D|S') = 0,01,$$

där S' betyder att personen är icke-sjuk, dvs frisk. Bayes formel ger

$$\mathbb{P}(S|D) = \frac{\mathbb{P}(D|S)\mathbb{P}(S)}{\mathbb{P}(D|S)\mathbb{P}(S) + \mathbb{P}(D|S')\mathbb{P}(S')} =$$

$$\frac{0,99 \times 10^{-4}}{0,99 \times 10^{-4} + 0,01 \times (1 - 10^{-4})} \approx 0,01.$$

(Konstigt? Tänk på en stad med 10.000 invånare, varav en är sjuk. Om alla testas, skulle vi vänta oss att den sjuka skulle upptäckas, och inte mindre än $0,01 \times 10000 = 100$ falsklarm.)

2. För definitionerna, se boken. Korrelationen är ett beroendemått, men egentligen mäter den graden av *linjärt* beroende. Därför medför inte nollkorrelation oberoende, medan motsatsen är fallet.

För uppgiftens andra del kan man utnyttja att kovariansen är linjär i sina bägge argument, om man vill göra det lätt för sig. Vill man arbeta lite mer kan man utgå från definitionen i termer av väntevärde.

3. Antag att mätningarna är oberoende och normalfördelade med samma väntevärde och varians. I så fall ges skattningen av deras aritmetiska medelvärde och konfidensintervallet bestäms ur t- fördelningen med tre frihetsgrader. Resultatet blir $1266 \pm 10,66$.

4. a. Sannolikheten att alla bokstäver är rätt blir $0,93^{10}$. sannolikheten för minst ett fel blir alltså ett minus detta tal, $1 - 0,93^{10} \approx 0,52$.

b. Enligt binomialfördelningen blir sannolikheten för exakt ett fel i själva meddelandet

$$\binom{10}{1} 0,01 \times 0,99^9 \approx 0,092,$$

om man kan förutsätta att felslagen är oberoende. Eftersom då antalet nollor ändras från udda till jämt eller tvärtom upptäcks felet om inte kontrolltecknet också blir fel. Bägge dessa ting inträffar samtidigt med sannolikheten $0,092 \times 0,01$.

5. a. Eftersom felet är normalfördelat sammanfaller trolighets-(maximum likelihood) och minsta kvadratskattningarna. Vi behöver alltså inte tveka om metodik, utom möjligen beträffande normeringen av den summa av kvadrater, som vi använder för att skatta felets varians, σ^2. Ska vi välja ML-skattningen

eller en justerad variant som blir väntevärdesriktig? (I rättningen har bägge svaren godkänts.)

I vårt fall förenklas bokens formler av att $\bar{x} = 0$, och skattningarna ges av

$$\alpha^* = \bar{y} \ (= 9,27),$$

$$\beta^* = \frac{\sum x_i y_i}{\sum x_i^2} \ (= 1,44),$$

$$\sigma^{2*} = \frac{\sum (y_i - \alpha^* - \beta^* x_i)^2}{n-2} \ (= 2,36), n = 11.$$

b. Intervallet har formen $\beta^* \pm t_{0,025}(9)\sigma^* / \sqrt{\sum x_i^2}$. Testet kan sedan utföras enligt sambandet mellan intervallskattning och hypotesprövning — förkasta om intervallet inte täcker 0.

c. Skattningen av y-värdet för $x = 3$ blir

$$y^*(3) = \alpha^* + \beta^* 3 \ (= 13,59).$$

Variansskattningen ges enligt boken av

$$\sigma^{2*} \left(\frac{1}{n} + \frac{n-2}{\sum x_i^2} \right),$$

och konfidensintervallet på sedvanligt sätt av t- fördelningen,

$$\alpha^* + \beta^* 3 \pm t_{0,025}(9)\sigma^{2*} \left(\frac{1}{n} + \frac{n-2}{\sum x_i^2} \right).$$

6. a. Eftersom Y är det minsta av $X_1, X_2, \ldots X_n$, och dessa är oberoende, blir

$$\mathbb{P}(Y > y) = \mathbb{P}(X_1 > y, X_2 > y, \ldots X_n > y) =$$

$$\mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) \ldots \mathbb{P}(X_n > y) = \left(\int_y^\infty e^{-(x-\theta)} \right)^n = e^{-n(y-\theta)},$$

förutsatt att $y \geq \theta$. Annars blir förstås $\mathbb{P}(Y > y) = 1$.Fördelningsfunktionen blir

$$F(y) = \begin{cases} 0 & \text{om } y \leq \theta, \\ 1 - e^{-n(y-\theta)} & \text{annars.} \end{cases}$$

Frekvensfunktionen fås på vanligt sätt som fördelningens derivata,

$$ne^{-n(y-\theta)}.$$

b.

$$\int_\theta^\infty yne^{-n(y-\theta)} = \theta + \int_\theta^\infty (y-\theta)ne^{-n(y-\theta)}$$

$$= \theta + \int_0^\infty yne^{-ny} = \theta + 1/n.$$

Tydligen ska vi välja $a = 1, b = -1/n$.

c. Och nu gäller det att räkna ut variansen. Men det är lätt nu när vi har frekvensfunktionen och väntevärdeskalkylen ovan.