Lösningar till Matematisk statistik för Z, del A den 17/1-02

1) Totalt finns $\binom{52}{5} = 2598960$ olika sätt att dra 5 kort. Antalet sätt att dra 3 av en specificerad valör (t ex dam) och 2 av en annan specificerad (t ex 7) är $\binom{4}{3}\cdot\binom{4}{2}\cdot\binom{44}{0} = 24$. Det finns 13 val av den valör vi ska dra 3 ifrån, och därefter 12 val av den valör vi ska dra 2 ifrån. Så den sökta sannolikheten är $\frac{13\cdot 12\cdot 24}{2598960} = \frac{6}{4165} \approx 0.00144058$.

2) Låt F betyda att området är förorenat och låt D betyda att mätningen detekterar förorening. Givet är att $P(F) = 0.5$, samt att $P(D|F) = 0.85$ och $P(D|F^c) = 0.3$. Bayes formel ger

$$P(F|D) = \frac{P(F)P(D|F)}{P(F)P(D|F) + P(F^c)P(D|F^c)} = \frac{0.5\cdot 0.85}{0.5\cdot 0.85 + 0.5\cdot 0.3} = 0.73913$$

3) T ex, om $Y = -X$, där $\mathrm{Var}[X] > 0$, så gäller att $\mathrm{Var}[X+Y] = 0$ (ty $X + Y = 0$), medan $\mathrm{Var}[X] + \mathrm{Var}[Y] = 2\mathrm{Var}[X] > 0$.

4) Låt N_t vara antalet impulser i intervallet $[0, t]$. Känt är att $N_t \sim \mathrm{Poi}(\lambda t)$. Ur ekvivalensen $W > t \Leftrightarrow N_t = 0$ följer nu att

$$P(W > t) = P(N_t = 0) = e^{-\lambda t}$$

och vi ser (t ex genom att konsultera formelsamlingen) att $W \sim \mathrm{Exp}(1/\lambda)$.

5) Det handlar i båda fallen om att integrera ut den andra variabeln. Vi får

$$f_X(x) = \int_0^x 6(1-x)\,dy = 6x(1-x)$$

för $0 < x < 1$ (då x är fixt, gäller ju att $0 < y < x$) och

$$f_Y(y) = \int_y^1 6(1-x)\,dx = \left(6x - \frac{6x^2}{2}\right)\Bigg|_{x=y}^{x=1} = (6-3) - (6y - 3y^2) = 3 - 6y + 3y^2 = 3(1-y)^2$$

för $0 < y < 1$ (då y är fixt gäller ju att $y < x < 1$).

6) Intervallskattningen $p = \hat{p} \pm 2\sqrt{\hat{p}(1-\hat{p})/n} = 0.293 \pm 0.029$ har den ungefärliga konfidensgraden 0.95 enl cgs. Se i formelsamlingen om skattning av proportioner.

7) Låt p vara sannolikheten som skattas vid det första tillfället och låt q den som skattas vid det senare. Obs att $\hat{q} = 608/2000 = 0.304$. Under $H_0 : p = q$ gäller att $p = q$ skattas av den sammanvägda proportionen $(293 + 608)/(1000 + 2000) = 0.300$, så teststatistikan får värdet

$$\frac{0.304 - 0.293}{\sqrt{0.300(1 - 0.300)(\frac{1}{1000} + \frac{1}{2000})}} = 0.620 \in (-1.96,\ 1.96)$$

Vi kan alltså inte förkasta $H_0 : p = q$ på nivån 5%. Se i formelsamlingen om jämförelse av proportioner.

8) Vi tänker oss att vi mäter θ med ett slumpmässigt fel ϵ, så att vårt mätvärde är $X = \theta + \epsilon$. Om vi nöjer oss med en mätning, är vårt resultat $X_1 = \theta + \epsilon_1$ och om vi gör två mätningar är det $(X_1 + X_2)/2 = \theta + (\epsilon_1 + \epsilon_2)/2$. Vi ska alltså jämföra felet från en mätning, vilket är ϵ_1 med felet $(\epsilon_1 + \epsilon_2)/2$ som vi har om vi gör två mätningar och bildar medelvärde. Vi som läst matematisk statistik vet att $\mathrm{Var}[(\epsilon_1 + \epsilon_2)/2] = (1/2)\mathrm{Var}[\epsilon_1]$ och att detta innebär att de statistiska variationer är mindre om vi gör två mätningar och bildar medelvärde. Ett sätt att förklara detta för en oinitierad skulle kunna vara att tänka sig in i det enkla fallet då felet ϵ bara kan anta värdena $\pm\delta$. Om vi inte ska ha något systematiskt fel, måste chansen vara lika stor, alltså 1/2 för båda värdena. Vi ser då att felet om vi gör två mätningar blir antingen $+\delta$,

0 eller $-\delta$. Så det blir iallafall inte större av medelvärdesbildningen. Men observera att chansen för $+\delta$ resp $-\delta$ är $1/4$ vardera, meden chansen att felet är 0 är $1/2$ p g a att de två felen tar ut varandra om det första är $+\delta$ och det andra $-\delta$ eller tvärtom. Vi har alltså

1 mätning

fel	chans
$+\delta$	$1/2$
$-\delta$	$1/2$

2 mätningar

fel	chans
$+\delta$	$1/4$
0	$1/2$
$-\delta$	$1/4$

Detta visar tydligt att det i detta enkla fall är bättre att göra två mätningar och medelvärdesbilda än att bara göra en.

Man kanske nu kan tänka sig in i fallet då felet ϵ har en fördelning (t ex normal) runt 0 och förklara att just det faktum att de resp felen kan ta ut varandra gör att (normal-) fördelningskurvan runt felet i det fall då man gör två mätningar och medelvärdesbildar blir smalare än i fallet då man bara gör en mätning.

Min tanke med uppgiften var att få fram att det är just det faktumet att oberoende fel har en tendens att ta ut varandra som gör medelvärdesbildning till en god idé. Det går självklart att förklara varför oberoende fel har en tendens att ta ut varandra på många bra sätt.