Statistik för Fysiker m. fl. Tentamen 12 januari 2004
Lösningar

Gräns för G 7p, för VG 17.

1. a. Läroboken ger beviset för det diskreta fallet, det kontinuerliga är helt analogt: Välj $g(X,Y) = X+Y$ i Sats 1', Kapitel 6 och utnyttja att en integral av en summa är summan av integralerna.

b. De två enklaste exempel jag kan tänka mig är $Y = X$ respektive $Y = -X$. Då blir i första fallet

$$V(X+Y) = V(2X) = 4V(X) > 2V(X),$$

om bara $V(X) > 0$, och i det andra

$$V(X+Y) = V(X-X) = V(0) < 2V(X).$$

c. Allmänt gäller ju att

$$V(X+Y) = V(X) + V(Y) + 2C(X,Y)$$

och kovariansen är noll om variablerna är oberoende (Sats 5, Kapitel 6, eller direkt härledning.)

2. a. $P(A|F) = P(A \cap F)/P(F)$, om $P(F) > 0$.

b. Tänk dig en situation som upprepas n gånger, varvid händelserna $A \cap F$ och F inträffar $n_{A\cap F}$ respektive n_F gånger. $P(A|F)$ bör vara ungefär frekvensen av A bland de gånger F inträffar, dvs

$$P(A|F) \approx n_{A\cap F}/n_F = \frac{n_{A\cap F}/n}{n_F/n} \approx P(A\cap F)/P(F).$$

c. Konstatera först att Bayes formel bara gäller om händelserna $F_0, F_1, \ldots F_n$ är oförenliga och tillsammans uttömmer alla möjligheter: $F_0 \cup F_1 \cup \ldots \cup F_n = \Omega$, dvs hela utfallsrummet.

I så fall är $P(F_0 \cap A) = P(A \cap F_0) = P(A|F_0)P(F_0)$, medan

$$P(A) = P(A \cap (F_0 \cup F_1 \cup \ldots \cup F_n)) = \sum_i P(A|F_i)P(F_i).$$

d. Bayes formel är viktig därför att den ger en möjlighet att gå från sannolikheten för en händelse (ovan betecknad A) under vissa betingelser (betecknade F) till sannolikheten att det var den betingelsen, som låg bakom A. T. ex. kan vi känna till sannolikheten för ett visst utslag på ett prov som mäter förekomsten av antikroppar, om patienten har en immunbristsjukdom. Har vi nu fått detta utslag kan vi använda formeln för att beräkna sanno-liokheten att sjukdomen faktikt föreligger.

3. a. Se boken.

b. Skriv B_n för händelsen A_n^c, dvs komplementet till A_n och $D_1 = B_1, D_2 = B_2 \cap A_1, \ldots D_n = B_n \cap A_{n-1} \ldots$. Kontrollera att D-händelserna är parvis oförenliga, att

$$B_n = \bigcup_{k=1}^{n} D_k$$

och att unionen av alla $D_1, D_2, \ldots$ blir hela utfallsrummet Ω. (Ett utfall som inte ligger i något D_k kan inte ligga i något B_n och måste därför tillhöra alla A_n. Men sådana utfall finns inte.) Därför blir

$$1 - P(A_n) = P(B_n) = \sum_{k=1}^{n} P(D_k) \to \sum_{k=1}^{\infty} P(D_k),$$

då $n \to \infty$. Men enligt bokens Axiom 3' är

$$\sum_{k=1}^{\infty} P(D_k) = P(D_1 \cup D_2 \cup \ldots D_k \cup \ldots) = P(\Omega) = 1.$$

4.a. Se boken, kapitel 6, exempel 4 (på två ställen)

b.

$$P([X] = k) = \int_k^{k+1} \frac{1}{m} e^{-x/m} dx = e^{-k/m} \int_0^1 \frac{1}{m} e^{-x/m} dx = e^{-k/m}(1 - e^{-1/m}).$$

c. Om $Y \in \mathrm{Exp}(m)$, så blir

$$\frac{f_Y(x)}{1 - F_Y(x)} = 1/m.$$

Om tvärtom

$$\frac{f_Y(x)}{1 - F_Y(x)} = c$$

och vi skriver $S(x) = 1 - F_Y(x)$, så gäller diferentialekvationen $S' = -cS$, med begynnelsevillkoret $S(0) = 0$. Det följer att

$$S(x) = e^{-cx}.$$

5.a. Man måste anta något om relationen mellan de två stokastiska variablerna. Tanken här är att de ska vara oberoende, när komponenterna betrakas var för sig. I så fall gäller att

$$P(\min(X, Y) > x) = P(X > x, Y > x) = e^{-ax}e^{-bx} = e^{-(a+b)x}.$$

Alltså är

$$F_{\min(X,Y)}(x) = 1 - P(\min(X, Y) > x) = 1 - e^{-(a+b)x}.$$

Detta gäller alltså om komponenterna *inte* påverkar varandra.

b. Men om de är smmankopplade gäller att båda går sönder, vid tidpunkten $\min(X, Y)$ med sannolikheten p. I övrigt är de oberoende. Alltså, om $x \leq y$, så

$$P(X \leq x, Y \leq y) = p(1 - e^{-(a+b)x}) + (1-p)(1 - e^{-ax})(1 - e^{-bx}).$$

6.a. Om $X_n = i$, så måste X_{n+1} vara antingen $i + 1$ eller $i - 1$, eftersom exakt en molekyl flyttar sig mellan behållarna i varje steg. Sannolikheten för den första möjligheten blir $1 - i/m$ och för den andra i/m, om $1 \leq i < m$. Om $i = 0$, kan X_{n+1} bara bli 1, och om $i = m$, måste nästa värdet i nästa steg bli $m - 1$. Sammanfattingsvis bestämmer X_n både vilka värden som är möjliga i nästa steg, och hur sannolika de är — det spelar ingen roll hur fördelningen mellan behållarna var tidigare.

b. Om molekylerna oberoende av varandra ligger i den ena behållaren med sannolikheten 1/2, blir

$$v_k = \binom{m}{k} 2^{-k} 2^{-(n-k)} = \binom{m}{k} 2^{-m}.$$

c. Vi måste visa att

$$v_k = v_{k-1}(1 - (k-1)/m) + v_{k+1}(k+1)/m$$

för $k = 1, 2, \ldots m-1$, $v_0 = v_1/m$, och $v_m = v_{m-1}/m$. Men det är bara att stoppa in

$$v_k = \binom{m}{k} 2^{-m} = \frac{m!}{k!(m-k)!} 2^{-m}$$

och kolla.

d. Vi vet att

$$P(X_{n+1} = k) = P(X_n = k-1)(1 - (k-1)/m) + P(X_n = k+1)(k+1)/m$$

Om $P(X_n = k) \to$ en gräns vi kan kalla π_k, då $n \to \infty$, så följer att talen $\pi_k, k = 0, 1, \ldots m$ måste uppfylla samma ekvation som de stationära sannolikheterna. Men denna ekvation har bara en lösning. Bara den stationära fördelningen är alltså tänkbar som gränsfördelning. (Det är svårare att visa att man verkligen har konvergens mot denna, men det erfordras inte.)

7. Se beskrivningen i läroboken.

8.a Eftersom $X+Y \in N(m, \sqrt{2})$, blir sannolikheten för att en individ, vars genomsnittliga (förväntade) tryck är m, skall kallas till vidareundersökning

$$P(X+Y > 20) = P(\frac{X+Y-m}{\sqrt{2}} > \frac{20-m}{\sqrt{2}}) = 1 - \Phi(\frac{20-m}{\sqrt{2}}).$$

b. På den första frågan blir svaret $1000\Phi((20-20)/\sqrt{2}) = 500$. Och på den andra $100.000(1 - \Phi((20-15)/\sqrt{2}) \approx 100.000(1 - \Phi(3.5355)) \approx 20$.

c. Kan man möjligen tänka sig att trycket i hela populationen (av 50-åringar) är normalfördelat, eller är det kloka kanske att föreslå en särskild undersökning av detta? Gränserna ovan verkar inte direkt väl valda. Vi skulle ju missa halva den sjuka populationen, och bara kalla ett ganska litet antal i onödan. Pröva dig fram med andra gränser, t. ex. 18 (då skulle vi missa ett åttiotal sjuka men kalla väldigt många till onödiga andraundersökningar).