Skriftlig Eksamen
Kombinatorik, sandsynlighed og randomiserede algoritmer (DM528)

Institut for Matematik & Datalogi
Syddansk Universitet

Torsdag den 7 Januar 2010, kl. 9–13

Alle sædvanlige hjælpemidler (lærebøger, notater etc.) samt brug af lommeregner er tilladt. Brug af computer er ikke tilladt.

Eksamenssættet består af 5 opgaver på 9 nummererede sider (1–9). Fuld besvarelse er besvarelse af alle 5 opgaver. De enkelte opgavers vægt ved bedømmelsen er angivet i procent. Der må gerne refereres til resultater fra lærebogen, samt noterne. Dette gælder også de opgaver der har været stillet til eksaminatorierne, eller til aflevering. Specielt må man gerne begrunde en påstand med at henvise til, at det umiddelbart følger fra et resultat i lærebogen, noterne, eller én af de opgaver, der har været stillet på ugesedlerne (hvis dette altså er sandt !). Henvisninger til andre bøger (ud over lærebogen) accepteres ikke som besvarelse af et spørgsmål! **Husk at begrunde dine svar!**

I Opgave 5 kan flere af delspørgsmålene besvares uafhængigt. Husk at du gerne må bruge det du skulle vise i et delspørgsmål i et senere spørgsmål også selvom du ikke har besvaret delspørgsmålet.

OPGAVE 1 (10%)

Bestem antallet af heltal n, hvor $1 \leq n \leq 999$ som ikke har nogen ens cifre (f.eks. er 2, 23 og 234 sådanne tal, medens 171 ikke er det).

OPGAVE 2 (15%)

Marie skal giftes i morgen ved en udendørs ceremoni i ørkenen. Hun har valgt stedet fordi det statistisk set kun regner 5 dage om året (som vi regner med har 365 dage). Desværre forudser den lokale meteorolog Bill regn i morgen. Om Bill's meteorologiske evner vides det, at når det bliver regnvejr, har Bill forudsagt det i 90% af tilfældene og når det ikke bliver regnvejr, så Bill alligevel forudsagt regn i 10 % af tilfældene. Bestem sandsynligheden for, at det bliver regnvejr på Maries bryllupsdag.

OPGAVE 3 (25%)

En lærer vil forhindre, at mange elever kan bestå hans multiple choice eksamen (to valg per spørgsmål) bare ved at gætte. Han laver derfor følgende regler:

1. Alle spørgsmål skal besvares.
2. Alle forkert besvarede spørgsmål tæller negativt. Dvs. hvis der er n_1 korrekte og n_2 forkerte, så får man $n_1 - n_2$ points samlet. Dvs. med n spørgsmål og $n_1 \leq n$ korrekte svar får man $2n_1 - n$ points. Man kan altså ende med negative points (helt ned til $-n$ points)!

Læreren benytter sin nye eksamensform til en eksamen med 10 spørgsmål. I denne eksamen deltager blandt andre eleverne A, B og C. A har jævnt styr på pensum og forventes at svare rigtigt på et vilkårligt spørgsmål i testen med sandsynlighed $\frac{3}{5}$. Elev B er bedre og svarer rigtigt på et vilkårligt spørgsmål i testen med sandsynlighed $\frac{4}{5}$. Elev C har ikke læst til eksamen og gætter bare på det ene af de to mulige svar i hvert spørgsmål. Han bruger en fair mønt til sine gæt, så han svarer korrekt på et vilkårligt spørgsmål i testen med sandsynlighed $\frac{1}{2}$.

Spørgsmål a:

Lad X_A, X_B, X_C værre stokastiske variable der betegner antallet af points som elev A, B, C opnår ved eksamen (husk at dette er antal rigtige minus antal forkerte). Bestem middelværdierne $E(X_A), E(X_B), E(X_C)$.

Spørgsmål b:

Bestem sandsynligheden for at $X_C > E(X_C)$.

Opgaven fortsættes!

Læreren har vedtaget, at man skal have mindst 2 points for at bestå eksamen.

Spørgsmål c:

Bestem sandsynligheden for at bestå, for hver af de tre studerende A, B, C.

Spørgsmål d:

Til eksamen er der 8 elever inklusive C som bare gætter. Hvor mange af disse må man forvente består eksamen?

Antag, at man anvender lærens eksamensmetode til tre multiple choice eksaminer med henholdsvis 100, 200 og 500 spørgsmål. Antag også, at kriteriet for at bestå er, at man får mindst 20% af de mulige points, når vi følger reglen om at trække antal forkerte fra antal rigtige.

Spørgsmål e:

Brug Chernoff bounds til at give en øvre grænse for sandsynlighederne for at en elev som C ville bestå hver af de tre eksaminer. Du skal forklare hvorfor du kan bruge Chernoff bounds. Passer de grænser du får med din intuition om hvor godt C vil klare sig når der er mange spørgsmål?

OPGAVE 4 (18%)

Bestem antallet af permutationer $(\pi(1), \pi(2), \pi(3), \pi(4), \pi(5))$ af tallene 1,2,3,4,5, som opfylder at $\pi(i+1) - \pi(i) \neq 2$ for $i = 1, 2, 3, 4$. Du kan foreksempel benytte inklusion-eksklusions princippet.

OPGAVE 5 (32%)

Spørgsmål a:

Bevis, at der for alle positive hele tal n, $n_1, n_2, \ldots, n_k$, hvor $2 \leq k \leq n$ og $\sum_{i=1}^{k} n_i = n$ gælder

$$\binom{n+1}{2} > \binom{n_1+1}{2} + \binom{n_2+1}{2} + \ldots + \binom{n_k+1}{2} \tag{1}$$

Du kan foreksempel anvende et kombinatorisk argument.

Resten af denne opgave går ud på at analysere en randomiseret algoritme til at konstruere en sammenhængende graf med n punkter.

Vi starter med at repetere lidt grafteori (du skal ikke vise noget af dette!). En graf med n punkter er sammenhængende hvis og kun hvis den har et udspændende træ; og ethvert sådant træ har $n-1$ kanter. Dette er også ækvivalent med at grafen har præcis én sammenhængskomponent. Hvis en graf har $k \geq 2$ sammenhængskomponenter $C_1, C_2, \ldots, C_k$ og vi tilføjer en ny kant, som går mellem to forskellige sammenhængskomponenter C_i og C_j, så har den nye graf $k-1$ sammenhængskomponenter, idet punkterne i $C_i \cup C_j$ nu tilhører den samme sammenhængskomponent. Hvis vi derimod tilføjer en kant uv hvor både u og v ligger i C_i for et eller andet i, så har den nye graf stadig k sammenhængskomponenter.

Lad $\mathcal{A}$ være følgende randomiserede algoritme. Algoritmens output er en sammenhængende graf $G = (V, E)$ med n punkter, som vi tænker på som en udspændende delgraf af den komplette graf $K_n = (\{1, 2, \ldots, n\}, \{ij : 1 \leq i < j \leq n\})$. **Husk** at K_n har $\binom{n}{2}$ kanter.

1. Sæt $E := \emptyset$, $V = \{1, 2, \ldots, n\}$
2. Sålænge $G = (V, E)$ har mindst 2 sammenhængskomponenter: vælg[1] en tilfældig kant $e = ij$ fra K_n og tilføj den til E hvis den ikke allerede er med i E.
3. Returner $G = (V, E)$.

[1] Vi vil også sige at $\mathcal{A}$ genererer en tilfældig kant fra K_n.

Bemærk, at $\mathcal{A}$ sagtens kan vælge den samme kant flere gange i forløbet og at dette naturligvis ikke fører til fremskridt, men stadig tæller med i det totale antal kanter som $\mathcal{A}$ når at vælge (generere).

Algoritmen $\mathcal{A}$ terminerer når der er er præcis 1 sammenhængskomponent i den aktuelle graf $G = (V, E)$, dvs. når $\mathcal{A}$ lige har tilføjet en kant uv mellem de entydigt bestemte forskellige sammenhængskomponenter i $G' = (V, E - \{uv\})$.

Spørgsmål b:

Gør rede for, at $\mathcal{A}$ skal lave mindst $n-1$ tilfældige kantvalg inden den terminerer og beskriv en situation hvor algoritmen kunne komme til at vælge $\Theta(n^2)$ forskellige kanter (dvs. ingen kant vælges mere end én gang) inden den terminerer.

For at analysere det forventede antal kanter som $\mathcal{A}$ samlet genererer inden den slutter, er det nyttigt at inddele forløbet af $\mathcal{A}$'s kørsel i n faser $0, 1, \ldots, n-1$, hvor vi siger at algoritmen er i fase i, hvis den aktuelle graf $G = (V, E)$ har præcis $n-i$ sammenhængskomponenter. Vi starter altså i fase 0 og algoritmen terminerer så snart fase $n-1$ nås. En anden måde at tænke på betydningen af de forskellige faser er, at når vi er i fase i har $\mathcal{A}$ indtil nu fundet præcis i kanter som gik imellem forskellige sammenhængskomponenter i den aktuelle graf da kanten blev genereret første gang (dvs vi har slået to sammenhængskomponenter sammen præcis i gange siden $\mathcal{A}$ startede).

Kald en kant uv **god** hvis u og v ligger i forskellige sammenhængskomponenter i den aktuelle graf $G = (V, E)$. Dvs i fase 0 er alle kanter gode, medens der ikke er noget gode kanter når fase $n-1$ er nået. Hvis en god kant uv vælges i fase i er denne fase slut, dvs. vi starter fase $i+1$, og vi kalder kanten uv **vigtig**, hvilket den forbliver i resten af algoritmen (vi starter altså med nul vigtige kanter og ender med præcis $n-1$ vigtige kanter som så udgør et udspændende træ i den endelige graf når $\mathcal{A}$ terminerer).

Lad os inddele de vigtige kanter i hver fase $i \geq 1$ efter de sammenhængskomponenter som de inducerer: lad $i_1, i_2, \ldots, i_k$, hvor $i = i_1 + i_2 + \ldots + i_k$ være antallet af vigtige kanter i de $k \geq 1$ sammenhængskomponenter af den aktuelle graf som har mere end et punkt (dvs mindst én vigtig kant i sig). Disse komponenter har så henholdsvis $i_1 + 1, i_2 + 2, \ldots i_k + 1$ punkter i sig og de resterende $n - i - k$ punkter fra G udgør hver deres egen sammenhængskomponent.

Spørgsmål c:

Gør rede for, at sandsynligheden for at vælge en god kant i fase i er mindst

$$\frac{\binom{n}{2} - \binom{i+1}{2}}{\binom{n}{2}}. \tag{2}$$

Lad X_i, $i = 0, 1, \ldots, n-2$ betegne antallet af tilfældige kanter som $\mathcal{A}$ genererer i fase i. Det vil sige $X = \sum_{i=0}^{n-2}$ er det totale antal kanter som $\mathcal{A}$ genererer i løbet af konstruktionen af G (inklusive dem der ligger inden i en sammenhængskomponent i den aktuelle graf og derfor ikke bidrager til et fase skift).

Spørgsmål d:

Gør rede for, at

$$E(X_i) \leq \frac{\binom{n}{2}}{\binom{n}{2} - \binom{i+1}{2}}. \tag{3}$$

Spørgsmål e:

Gør rede for, at

$$E(X) \leq \binom{n}{2} \sum_{i=0}^{n-2} \frac{1}{\binom{n}{2} - \binom{i+1}{2}}. \tag{4}$$

Spørgsmål f:

Gør rede for, at der for alle heltal i, n med $0 \leq i \leq n-2$ gælder at

$$\left(\frac{n-1}{2}\right) \left(\frac{1}{\binom{n}{2} - \binom{i+1}{2}}\right) \leq \frac{1}{n-i}. \tag{5}$$

Spørgsmål g:

Benyt (4) og (5) til at vise at

$$E(X) = \mathcal{O}(n \ln n). \tag{6}$$

Diskuter hvad dette siger om algoritmen $\mathcal{A}$.