ECON 2130

OBLIGATORISK SEMESTEROPPGAVE 2008 VÅR

Innlevering: Mandag 14. april (SV's infosenter i 1. etg. ES). En standardisert førsteside som ligger på nettet, skal fylles ut og legges ved oppgaven.

Merk: Det er lov å samarbeide, men hver skal levere egen rapport. Plagiater vil ikke bli godkjent (dette gjelder også den som har latt en annen skrive av sin besvarelse). Det er ikke meningen at alt skal løses på PC. Bruk PC der det er hensiktsmessig. Merk også at vedlagte PC utskrifter bør redigeres og kommenteres. Ta ikke med alt som kommer ut av maskinen, bare det som er av betydning/interesse for besvarelsen! (Det er for eksempel bortkastet å legge ved en utskrift av de 10000 observasjonene som du skal generere i oppgaven). En besvarelse som bare består av en bunke ukommenterte utskrifter blir høyst sannsynlig ikke godkjent.

Oppgave

Vi tar utgangspunkt i følgende eksperiment: En rettferdig terning kastes tre ganger. La X være antall seksere som oppnås og Y antall enere. Begge variablene har dermed verdimengden $\{0,1,2,3\}$. Det kan vises (du trenger ikke å gjøre det her) at simultanfordelingen for (X,Y) er gitt ved følgende formel

(1) $$f(x,y) = P(X = x \cap Y = y) = \frac{3!}{x!\,y!(3-x-y)!} \cdot \left(\frac{1}{6}\right)^{x+y} \cdot \left(\frac{4}{6}\right)^{3-x-y}$$

for $x, y = 0,1,2,3$ og slik at $0 \le x + y \le 3$.

A. Skriver vi ut fordelingen i en tabell, får vi

Tabell 1. **(** $f(x,y)$ **)**

	y = 0	y = 1	y = 2	y = 3	Sum
x = 0	8/27	2/9	1/18	1/216	125/216
x = 1	2/9	1/9	1/72	0	75/216
x = 2	1/18	1/72	0	0	15/216
x = 3	1/216	0	0	0	1/216
Sum	125/216	75/216	15/216	1/216	1

Verifiser ved innsetting i formelen (1) at $f(0,0) = 8/27$ og $f(2,1) = 1/72$. Forklar hvorfor *X* og *Y*, hver for seg, er binomisk fordelte. Sett opp formler for de to fordelingene og verifiser at de faller sammen med marginalfordelingene i tabell 1.

B. Finn E(*X*), E(*Y*), var(*X*), var(*Y*). Verifiser at $\text{cov}(X,Y) = -1/12$ og korrelasjonen, $\text{corre}(X,Y) = -1/5$.

C. Følgende spill blir tilbudt. Mot å betale 100 kr. i spilleavgift, kan spilleren kaste en (rettferdig) terning tre ganger. For hver sekser som oppnås får spilleren utbetalt 200 kr., men må (i tillegg til spilleavgiften) betale 20 kr for hver ener som måtte dukke opp. Uttrykt ved *X* og *Y*, blir dermed gevinsten (*V*) for spilleren $V = 200X - 20Y - 100$ for et spill. Fortjenesten for et spill *for tilbyderen* er $-V$ kr.

Verifiser at $\text{E}(V) = -10$ og $\text{var}(V) = 17500$ [**Hint:** Bruk for eksempel regel 4.15 i boka (regel 119 i utg. 1)]. Hva blir forventningen og variansen til fortjenesten (for et spill) for tilbyderen?

.

D. Finn sannsynligheten for at V er større enn 50 kr. [**Hint:** Det kan, f.eks., være lurt å lage en tabell, som tabell 1, som viser verdien av *V* for alle mulige kombinasjoner av *X* og *Y*, og deretter utnytte tabell 1.]

E. Finn fordelingen til *V*. Det vi er ute etter er en tabell av formen

v	v_1	v_2	$\cdots$	v_k
$h(v) = P(V = v)$	$h(v_1)$	$h(v_2)$	$\cdots$	$h(v_k)$

der første rad gir verdimengden for *V*, og annen rad de tilhørende sannsynlighetene.

Lag et histogram over fordelingen. (Hvis du ikke får Excel til å gi noe tilfredsstillende histogram, skisser et for hånd.)

F. La *U* være være total gevinst etter 20 spill. Dvs. la $V_1, V_2, \ldots, V_{20}$ være uavhengige og identisk fordelte som *V*. Da er $U = V_1 + V_2 + \cdots + V_{20}$. Finn forventningen og standardavviket til *U*. Nevn hvilken eller hvilke regler i boka du bruker til dette.

Vi er interessert i sannsynligheten for at tilbyderen får positiv fortjeneste på 20 spill, med andre ord $P(U < 0)$. Den eksakte fordelingen til *U* er meget komplisert, men i følge sentralgrenseteoremet (jfr. regel 5.18 i boka), kan vi regne *U* som tilnærmet normalfordelt. Bruk dette til å beregne $P(U < 0)$ tilnærmet.

G. I boka er det angitt en tommelfingerregel at antall ledd i summen (*U*) bør være minst 20 for at tilnærmelsen til normalfordelingen skal være tilfredstillende. Vår *U* er basert på 20 spill og representerer dermed et grensetilfelle. Du skal nå bruke datamaskinen til å simulere 500 observasjoner av *U*. For hver observasjon av *U* trenger du 20 uavhengige observasjoner av *V*, som summeres for å gi *U*. I Excel er det mulig å trekke observasjoner fra fordelingen til *V* i punkt **E**. Du lager først to kolonner ved siden av hverandre, der verdimengden for *V* utgjør første kolonne og de tilhørende sannsynlighetene den andre kolonnen. Deretter går du inn in ”Random Number Generation” under ”Data Analysis” under ”Tools”. I menyen som kommer fram velger du *Discrete* som fordeling, Number of Variables = 20, Number of Random Numbers = 500. Deretter markerer du området der fordelingen for *V* står, i vinduet for ”Value and Probability Input Range”, og, til slutt, den første cellen i ”Output Range”.

Du har nå generert (simulert) 500x20=10000 uavhengige observasjoner av *V*, organisert i 20 kolonner. Hver rad representerer et observasjonssett av $V_1, V_2, \ldots, V_{20}$. Beregn summen av hver rad i en ny kolonne, og du har oppnådd 500 uavhengige observasjoner av *U*. Finn gjennomsnitt og empirisk standardavvik for *U*-observasjonene og sammenlign med de teoretiske størrelsene, $\mathrm{E}(U)$, $\mathrm{SD}(U)$. Lag et histogram for *U*-ene basert på ca. 15 intervaller (finn minimum og maksimum for *U*-ene og velg for eksempel 15 passende intervallgrenser (”bins”), som du legger inn i en kolonne i Excel osv.).

H. Lag også et histogram for *U*-observasjonene som figur 5.23 i boka (også i utg. 1) der tettheten for den tilnærmende normalfordelingen fra sentralgrenseteoremet er tegnet inn. Dette er vanskelig å få til i Excel, så tegn et for hånd på rutepapir. Husk å velge riktig skala på Y-aksen. Virker tilnærmelsen til normalfordelingen rimelig? [**Hint:** Funksjonsverdier for den aktuelle normalfordelingstettheten kan du enklest beregne med NORMDIST-funksjonen i Excel som du finner under *statistical functions*.]

I. Finn den relative frekvensen av *U*-observasjoner som er negative (dvs. som gir positiv fortjeneste for tilbyderen). I følge de store talls lov vil dette tallet ligge i nærheten av $P(U < 0)$ og kan derfor ses på som et anslag (estimat) på den sanne verdien av $P(U < 0)$. Sammenlign den relative frekvensen med tilnærmingen til den teoretiske verdien av $P(U < 0)$ beregnet i punkt **F**. Gir dette, syns du, grunnlag for å tvile på tilnærmingen foretatt i punkt **F**? [**Hint:** Du trenger å telle opp hvor mange av *U*-observasjonene er negative. Dette kan du, for eksempel, la Excel gjøre på følgende måte: Lag en ny kolonne ved siden av *U*-kolonnen med 0-er og 1-ere slik at det står et 1-tall ved siden av *U*-er som negative og 0 ved siden av de andre. Ved å summere denne kolonnen får du antall negative *U*-er. Kolonnen kan lages med IF-funksjonen. For eksempel, hvis første *U*-tall står i celle ”X2”, skriv i cellen ved siden av, =IF(X2<0;1;0). (Merk at det er semikolon i formelen – ikke komma.) Kopier deretter denne cellen til resten av kolonnen.]

J. Beregn $P(U < 0)$ der U nå utgjør totalgevinsten for 500 spill ($U = V_1 + V_2 + \cdots + V_{500}$). Du kan nå, uten tap av realisme, regne at U er normalfordelt. Finn også det mest sannsynlige variasjonsområdet for U, i følgende forstand: Bestem c_1 og c_2 slik at $P(c_1 < U < c_2) = 0,90$. Forklar hvorfor det gir riktig svar hvis vi bestemmer c_1 slik at $P(U \leq c_1) = 0,05$ og c_2 slik at $P(U \leq c_2) = 0,95$.