Om fordelingen til $X + Y$

STK1100 V11

1. Variansen til $X + Y$
2. Direkte utledninger av fordelingen til $X + Y$
3. Bruk av momentgenererende funksjoner
4. Konvolusjonsformelen

Om fordelingen til $X + Y$

Det er i en rekke situasjoner av interesse å finne fordelingen til $X + Y$ når vi kjenner fordelingene til X og Y.

Vi skal se at mange fordelinger er "lukket" med respekt til summer, dvs. $X + Y$ har samme type fordelingen som X og Y

Dette er også et første skritt mot å finne fordelingen til gjennomsnittet

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

selv om vi for gjennomsnittet skal se at vi kan tilnærme med en normalfordeling.

Variansen til $X + Y$

Vi viste at generelt, dvs. også når X og Y er avhengige gjelder

$$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$$

Med μ_X og μ_Y forventningen til X og Y har vi da

$$\begin{aligned}
V(X+Y) &= \mathrm{E}[\{(X+Y) - (\mu_X + \mu_Y)\}^2] \\
&= \mathrm{E}[\{(X-\mu_X) + (Y-\mu_Y)\}^2] \\
&= \mathrm{E}[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_x)(Y-\mu_Y)] \\
&= \mathrm{E}[(X-\mu_X)^2] + \mathrm{E}[(Y-\mu_Y)^2] \\
&\quad +2\mathrm{E}[(X-\mu_x)(Y-\mu_Y)] \\
&= V(X) + V(Y) + 2\mathrm{Cov}(X,Y)
\end{aligned}$$

som altså gjelder generelt for avhengige stokastiske variable X og Y (med eksisterende varianser).

Variansen til $X + Y$, forts.

Spesielt hvis $\mathrm{Cov}(X, Y) = 0$ får vi da

$$V(X + Y) = V(X) + V(Y)$$

Det ble vist at hvis X og Y er uavhengige stokastiske variable så gjelder nettopp

$$\mathrm{Cov}(X, Y) = 0$$

og vi har da at

Variansen til en sum = Summen av variansene

Fordelingen til $X + Y$, Generelt

Vi kan finne fordelingen til $X + Y$ ved 3 typer framgangsmåter

- Direkte argument
- Ved hjelp av momentgenererende funksjoner
- Ved "konvolusjonsformelen"

Fordelingen til $X + Y$, Generelt

- Et direkte argument kan bare brukes i noen tilfeller
- momentgenererende funksjoner kan benyttes i noen ytterligere situasjoner
- Begge disse metodene vil kreve uavhengige X og Y
- Konvolusjonsformelen gir en generell metode for å generere fordelingen til $X + Y$

Eksempel: Direkte argument

Anta at $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ og X og Y uavhengige, dvs. X og Y binomiske med samme suksess-sannsynlighet p, men muligens ulike n og m.
Vi har da

$$\begin{aligned} X &= \text{ antall suksesser i } n \text{ forsøk} \\ Y &= \text{ antall suksesser i } m \text{ forsøk} \end{aligned}$$

der forsøkene er uavhengige med like suksess-sannsynlighet p (Bernoulli-sekvens). Men da er jo også

$$X + Y = \text{ antall suksesser i } n + m \text{ forsøk}$$

i en Bernoulli-sekvens på $n + m$ forsøk. Altså

$$X + Y \sim \text{Bin}(n + m, p)$$

$X + Y$ ved momentgenererende funksjoner

Vi skal utlede det samme resultatet ved hjelp av momentgenererende funksjoner $M_X(t) = \text{E}[e^{Xt}]$.
Husk at $M_X(t)$ identifiserer fordelingen til X.

Vi skal bruke følgende resultat

Proposisjon Anta at X og Y er uavhengige med momentgenererende funksjoner $M_X(t)$ og $M_Y(t)$. Da holder

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

Momentgenererende funksjon for binomiske X og Y

Anta igjen at $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ og X og Y uavhengige. Da har (vist tidligere) X mom.gen. funk.

$$M_X(t) = [1 - p + pe^t]^n$$

og tilsvarende for Y. I henhold til Proposisjonen får da $X + Y$ mom.gen. funk.

$$\begin{aligned} M_{X+Y}(t) = M_X(t)M_Y(t) &= [1 - p + pe^t]^n [1 - p + pe^t]^m \\ &= [1 - p + pe^t]^{n+m} \end{aligned}$$

Men dette er den mom.gen. funk. for $\text{Bin}(n + m, p)$, dvs. for en binomisk variabel med $n + m$ forsøk og suksess-sannsynlighet p, altså

$$X + Y \sim \text{Bin}(n + m, p)$$

Bevis for proposisjon

Siden X og Y er uavhengige har de simultantetthet

$$f(x, y) = f_X(x)f_Y(y)$$

dvs. som produktet av marginaltetthetene. Da får vi, med dobbelt-integral over $\Re^2$, for funksjoner $g(x)$ og $h(y)$,

$$\begin{aligned} \text{E}[g(X)h(Y)] &= \int\int g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= [\int g(x)f_X(x)dx][\int h(y)f_Y(y)dy] \\ &= \text{E}[g(X)]\text{E}[h(Y)] \end{aligned}$$

Dermed blir, siden X og Y er uavhengige,

$$M_{X+Y}(t) = \text{E}[e^{t(X+Y)}] = \text{E}[e^{tX}e^{tY}] = \text{E}[e^{tX}]\text{E}[e^{tY}] = M_X(t)M_Y(t)$$

Når fører et direkte argument fram?

Det er ihvertfall tre situasjoner hvor et direkte probabilistisk argument kan brukes

- Binomisk fordeling (som vist)
- Negativt binomiske (og geometriske) fordelinger med samme p
- Poisson-fordelinger: Anta X og Y uavhengige med $X \sim \text{Po}(\lambda), Y \sim \text{Po}(\mu)$. Da er $X + Y \sim \text{Po}(\lambda + \mu)$.

For Poisson-fordeling benyttes sammenhengen med Poisson-prosess.

Oppgave: Benytt et direkte argument for å finne fordelingen til $X + Y$ i de to siste tilfellene.

Når fører argument basert på mom.gen. funk. fram?

- Binomisk (som vist)
- Neg. bin. (og geometriske) fordelinger med samme p
- Poisson-fordelinger
- Eksponensial-fordelinger med samme λ
- Gamma-fordelinger med samme β
- Normalfordelinger

Selvfølgelig må X og Y være uavhengige.

Oppgave Utled disse resultatene.

X og Y normalfordelte

Anta $X \sim \mathrm{N}(\mu_X, \sigma_X)$ og $Y \sim \mathrm{N}(\mu_Y, \sigma_Y)$. Da er

$$M_X(t) = e^{\mu_X t + \sigma_X^2 t^2/2}$$

og tilsvarende for Y. Dermed blir momentgenererende for $X+Y$

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) = e^{\mu_X t + \sigma_X^2 t^2/2} e^{\mu_Y t + \sigma_Y^2 t^2/2} \\ &= e^{(\mu_X + \mu_Y)t + (\sigma_X^2 + \sigma_Y^2) t^2/2} \end{aligned}$$

som er mom.gen. funksjon for $\mathrm{N}(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$, dvs. $X+Y$ har en normalfordeling med forventning $\mu_X + \mu_Y$ og varians $\sigma_X^2 + \sigma_Y^2$.

Konvolusjonsformelen, diskret fordelinger

Anta at X og Y er diskrete stokastiske variable med simultane pkt.sh $p(x,y)$ og la $U = X+Y$. Vi kan skrive begivenheten

$$\{U = u\} = \bigcup_{\text{alle } x} \{X = x, Y = u - x\}$$

som er en disjunkt union. Dermed får vi

$$\mathrm{P}(U = u) = \sum_{\text{alle } x} \mathrm{P}\{X = x, Y = u - x\} = \sum_{\text{alle } x} p(x, u - x)$$

Spesielt hvis X og Y er uavhengige og $p(x,y) = p_X(x) p_Y(y)$ fås

$$\mathrm{P}(U = u) = \sum_{\text{alle } x} p_X(x) p_Y(u - x)$$

Dette kalles **konvolusjonsformelen**.

Summen av to Poisson-fordelte variable

Anta $X \sim \mathrm{Po}(\lambda)$, $Y \sim \mathrm{Po}(\mu)$ og X og Y uavhengige.
Da blir ved konvolusjonsformelen, for $U = X+Y$,

$$\mathrm{P}(U = u) = \sum_{\text{alle } x} p_X(x) p_Y(u - x) = \sum_{x=0}^{u} p_X(x) p_Y(u - x)$$

siden $p_X(x) p_Y(u-x) > 0$ når $x = 0, 1, 2, \ldots, u$. Dermed

$$\begin{aligned} \mathrm{P}(U = u) &= \textstyle\sum_{x=0}^{u} \frac{\lambda^x}{x!} e^{-\lambda} \frac{\mu^{u-x}}{(u-x)!} e^{-\mu} \\ &= \frac{1}{u!} e^{-(\lambda+\mu)} \textstyle\sum_{x=0}^{u} \frac{u!}{x!(u-x)!} \lambda^x \mu^{u-x} \\ &= \frac{1}{u!} e^{-(\lambda+\mu)} (\lambda + \mu)^u \end{aligned}$$

der den siste likheten følger av Newton's binomialformel.
Men dette er nettopp en Poisson-sannsynlighet med forventning $\lambda + \mu$, dvs. $U = X + Y \sim \mathrm{Po}(\lambda + \mu)$.

Konvolusjonsformelen for kontinuerlige stok. var.

Tilsvarende det diskrete tilfelle vil vi ha for (X,Y) kontinuerlig fordelt med simultantetthet $f(x,y)$ at tettheten til $U = X+Y$ gis ved

$$f_U(u) = \int_{-\infty}^{\infty} f(x, u - x) dx$$

Spesielt når X og Y er uavhengige med tettheter $f_X(x)$ og $f_Y(y)$ fås konvolusjonsformelen

$$f_U(u) = \int_{-\infty}^{\infty} f_X(x) f_Y(u - x) dx$$

Eksempel: X og Y uavhengige og $\mathrm{N}(0,1)$

Når X og Y begge er standard normalfordelte gir konvolusjonsformelen at $U = X + Y$ har tettet

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u-x)^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-u^2/4} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-2(x-u/2)^2/2} \sqrt{2} dx \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-u^2/4} \end{aligned}$$

siden $x^2 + (u-x)^2 = \frac{u^2}{2} + 2(x - \frac{u}{2})^2$ og det siste integralet blir lik 1 (er over en tetthet).

Men dette betyr at $U \sim \mathrm{N}(0, \sqrt{2})$, normalfordelt med forventning 0 og varians lik 2.

Noe som også følger av argumentet med mom.gen.fu.

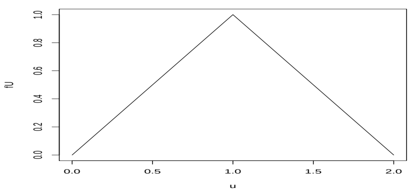
Fordelingen til summen av to $U[0,1]$-variable

Anta at både X og Y er uavhengige og uniformt fordelt på $[0,1]$, dvs. begge har tetthet

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{ellers} \end{cases}$$

Fordelingen til $U = X + Y$ blir da

$$f_U(u) = \int_0^1 f(x) f(u-x) dx = \begin{cases} \int_0^u 1 dx = u & 0 < u < 1 \\ \int_{u-1}^1 1 dx = 2 - u & 1 < u < 2 \end{cases}$$



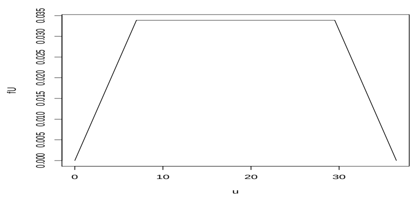
"Påskefordeling" (?)

La U = tid fra vårjevndøgn til 1. påskedag $= X + Y$ der

- X = tid fra vårjevndøgn til første fullmåne
- Y = tid fra fullmåne til første søndag

Rimelig å anta $X \sim U[0, 29.5]$ og $Y \sim U[0,7]$. Antar også at X og Y er uavhengige. Får

$$f_U(u) = \int_0^{29.5} f_X(x) f_Y(u-x) dx = \begin{cases} \frac{u}{7*29.5} & 0 < u < 7 \\ \frac{1}{29.5} & 7 < u < 29.5 \\ \frac{1}{29.5} - \frac{u-29.5}{7*29.5} & 29.5 < u < 36. \end{cases}$$



1. påskedag 1700-2299