Gammafordelingen og χ^2-fordelingen

Gammafunksjonen

Gammafunksjonen er en funksjon som brukes ofte i sannsynlighetsregning. I mange fordelinger dukker den opp i konstantleddet.

Hvis man plotter n-fakultet ($n!$) mot n for $n = 0, 1, \ldots$ er det naturlig å spørre seg om det ikke finnes en kontinuerlig funksjon som går gjennom alle punktene. Svaret på dette spørsmålet er nettopp gammafunksjonen, eller rettere sagt, en translasjon av gammafunksjonen. Mer presist, den funksjonen vi skal definere og skrive som $\Gamma(n)$ vil for heltallige n være lik $(n-1)!$. Altså er $\Gamma(1) = 0! = 1$, $\Gamma(2) = 1! = 1$, $\Gamma(3) = 2! = 2$, osv.

Vi definerer gammafunksjonen for alle positive x ved integralet

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt. \tag{1}$$

Vi ser da at

$$\Gamma(1) = \int_0^\infty e^{-t} dt = [-e^{-t}]_0^\infty = 0 - (-1) = 1.$$

Ved delvis integrasjon finner vi videre at

$$\begin{aligned} x\Gamma(x) &= \int_0^\infty x t^{x-1} \cdot e^{-t} dt = \int_0^\infty u'(t)v(t) dt = [t^x e^{-t}]_0^\infty - \int_0^\infty t^x(-e^{-t}) dt \\ &= 0 + \int_0^\infty t^{(x+1)-1} e^{-t} dt = \Gamma(x+1). \end{aligned}$$

Dermed gjelder rekursjonsformelen

$$\Gamma(x+1) = x\Gamma(x) \tag{2}$$

generelt. Siden $\Gamma(1) = 1$ finner vi da rekursivt $\Gamma(2) = 1 \cdot \Gamma(1) = 1$, $\Gamma(3) = 2\Gamma(2) = 2!$, $\Gamma(4) = 3\Gamma(3) = 3!$, ... osv, slik at $\Gamma(n) = (n-1)!$.

Det kan vises at $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Ved å bruke likning (2) kan man da beregne funksjonen for alle halvtallige verdier.

For andre verdier må funksjonsverdien beregnes numerisk, enten ved å benytte en rekkeutvikling for den eller ved numerisk integrasjon.

Oppgave 1

Hvordan oppfører $\Gamma(x)$ seg når x nærmer seg null (ovenfra)? Bruk rekursjonsformelen for gammafunksjonen til å beregne $\Gamma(0.5), \Gamma(1.0), \Gamma(1.5), \ldots \Gamma(4)$. Bruk disse verdiene til å skissere gammafunsjonen på området $[0, 4]$.

Gammafordelingen

Vi kan nå finne konstanten c i fordelingen på formen

$$f(x) = cx^{k-1}e^{-\alpha x}$$

definert for positive x. Ved å innføre ny integrasjonsvariabel $u = \alpha x$ finner vi

$$\int_0^\infty f(x)dx = c\int_0^\infty (\frac{u}{\alpha})^{k-1}e^{-u}\frac{1}{\alpha}du = \frac{c}{\alpha^k}\int_0^\infty u^{k-1}e^{-u}du = \frac{c\Gamma(k)}{\alpha^k}.$$

Siden dette integralet må være lik 1 for at $f(x)$ skal være en sannsynlighetstetthet ser vi at konstanten c må være lik $\alpha^k/\Gamma(k)$. Den fordelingen vi nå har funnet kalles *gammafordelingen.*

Vi sier at X er gammafordelt med *formparameter* k og *skalaparameter* α hvis X har sannsynlighetstettheten

$$f(x) = \frac{\alpha^k}{\Gamma(k)}x^{k-1}e^{-\alpha x} \tag{3}$$

for $x \geq 0$.

Legg merke til at eksponensialfordelingen er en gammafordeling med formparameter lik 1.

Vi kan nå også finne forventningen og variansen til X ved å beregne

$$\mathrm{E}X^r = \int_0^\infty x^r f(x)dx = \frac{\alpha^k}{\Gamma(k)}\int_0^\infty x^{(k+r)-1}e^{-\alpha x}dx.$$

Ved å erstatte k med $(k{+}r)$ i resultatet funnet ovenfor, som sier at $\int x^{k-1}e^{\alpha x}dx = \Gamma(k)/\alpha^k$, finner vi

$$\mathrm{E}X^r = \frac{\alpha^k}{\Gamma(k)} \frac{\Gamma(k+r)}{\alpha^{k+r}} = \frac{\Gamma(k+r)}{\alpha^r\Gamma(k)}.$$

Ved å sette inn $r = 1$ og bruke at $k\Gamma(k) = \Gamma(k+1)$ gir dette

$$\mathrm{E}X = \frac{k}{\alpha}$$

og tilsvarende, ved å sette $r = 2$ får vi at $\mathrm{E}X^2 = k(k+1)/\alpha^2$ som gir

$$\mathrm{var}(X) = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{k}{\alpha^2}.$$

Oppgave 2

Skjevheten til en fordeling med forventning μ og varians σ^2 er definert som $\mathrm{E}(X-\mu)^3/\sigma^3$. Symmetriske fordelinger (f. eks. normalfordelingen) vil da ha skjevhet lik null. Finn skjevheten til gammafordelingen og studer hvordan den forandrer seg når formparameteren forandrer seg.

Hint: multipliser ut $(X-\mu)^3$ og finn alle ledd ved å sette inn i den generelle formelen for momentene til gammafordelingen. Bruk rekursjonsformelen til å bli kvitt gammafunksjonen.

Skalatransformasjoner og summer

La nå X være gammafordelt med parametre (k, α) og definer $Y = \beta X$, for $\beta > 0$. Ved å benytte formelen for monotone transformasjoner finner vi da at Y også er gammafordelt. Formparameteren blir den samme som for X, dvs. lik k, mens skalaparameteren blir α/β.

En annen nyttig egenskap ved gammafordelingen er følgende: Dersom $X_1, X_2, \ldots, X_n$ er uavhengige variable, X_i gammafordelt med formparameter k_i og skalaparameter α, så er summen $Z = \sum_{i=1}^n X_i$ også gammafordelt. Formparameteren i fordelingen til Z er $\sum_{i=1}^n k_i$ og skalaparameteren er α. Av resultatet om endring av skala gitt ovenfor følger det da at middelverdien Z/n også blir gammafordelt. Formparameteren blir $\sum k_i$ og skalaparameteren αn.

Hvis vi adderer mange variable med samme gammafordeling vil summen være gammafordelt med stor formparameter. Ifølge sentralgrenseteoremet kan derfor gammafordelingen tilnærmes med en normalfordeling når formparameteren er stor. I praksis fungerer dette godt når $k > 10$.

Oppgave 3

La X være gammafordelt med parametre (k, α). Bruk transformasjonsformelen for monotone transformasjoner til å vise at $Y = \beta X$ da er gammafordelt med parametre $(k, \alpha/\beta)$.

χ^2-fordelingen

I statistisk analyse av normalfordelte variable anvendes en fordeling som er et spesialtilfelle av gammafordelingen og som kalles χ^2-fordelingen.

La X være en standard normalfordelt variabel og sett $Y = X^2$. Vi kan da finne den kumulative fordelingen til Y,

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Vi deriverer på begge sider m.h.p. y og finner sannsynlighetstettheten

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})].$$

Ved å sette inn i standard normalfordelingen og bruke at $\sqrt{\pi} = \Gamma(\frac{1}{2})$ følger det at at

$$f_Y(y) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} y^{1/2-1} e^{-y/2}$$

for $y > 0$. Vi ser at $Y = X^2$ er gammafordelt med formparameter og skalaparameter lik $1/2$.

La nå $X_1, X_2, \ldots, X_n$ være uavhengige standard normalfordelte variabler og sett $Y = X_1^2 + X_2^2 + \ldots + X_n^2$. Da er Y en sum av n gammafordelte variable med parametre $(1/2, 1/2)$ og er derfor gammafordelt med parametre $(n/2, 1/2)$, dvs.

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-n/2}. \tag{4}$$

Denne fordelingen kalles χ^2-fordelingen med n frihetsgrader. Begrepet 'frihetsgrader' brukes her som en betegnelse på en heltallig parameter. Ved å sette inn i uttrykkene for forventning og varians til gammafordelingen finner vi at forventningsverdien blir n og variansen $2n$. Kvantiler i denne fordelingen for gitte verdier av n finner man i statistiske tabeller.

Vi ser at hvis n er stor, så er Y en sum av et stort antall variable (X_i^2) og er dermed tilnærmet normalfordelt ifølge sentralgrenseteoremet. Tilnærmelsen vil være ganske nøyaktig for $n > 20$.

Oppgave 4

Finn 5%- og 95%-kvantilene i χ^2-kvadratfordelingen med hhv. 10, 20 og 50 frihetsgrader i en statistisk tabell. Beregn også disse kvantilene ved å tilnærme fordelingene med normalfordelinger og sammenlikn med de eksakte tallene.

Oppgave 5

La X være gammafordelt med parametre (k, α). Vis at fordelingen til $Y = 2\alpha X$ er χ^2-fordelt, men at 'antall frihetsgrader' ikke nødevendigvis er et helt tall. Bruk dette til å finne tilnærmet 5%-kvantilen i gammafordelingen med parametre $(7.3, 3.4)$.

Hint: Interpoler i χ^2-tabellen.

Bruk av χ^2-fordelingen i statistisk analyse

La nå $X_1, X_2, \ldots, X_n$ være uavhengige $N(\mu, \sigma^2)$. Da er $(X_i - \mu)/\sigma$ standard normalfordelte og det følger av resultatene ovenfor at kvadratsummen

$$Z = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

er χ^2-fordelt med n frihetsgrader.

Et resultat som viser å være mer anvendbart er at vi fremdeles får en χ^2-fordelt variabel dersom forventningen μ erstattes med middelverdien $\bar{X}$, men at antall frihetsgrader da er $n-1$. Dette resultatet vil blant annet bli brukt til å trekke slutninger om en ukjent varians σ^2 på grunnlag av uavhengige normalfordelte observasjoner.

Et annet nyttig resultat er at summer (og middelverdier) av eksponensielt fordelte variable også kan analyseres ved hjelp av χ^2-fordelingen. Hvis X_i, $i = 1, 2 \ldots, n$, er uavhengige eksponensielt fordelte variable med sannsynlighetstetthet

$$f(x) = \lambda e^{-\lambda x}$$

så følger det av resultatene ovenfor at $\sum_{i=1}^{n} X_i$ er gammafordelt med formparameter n og skalaparameter λ. Videre følger det av resultatet for skalaendringer at $2\lambda \sum X_i$ da er gammafordelt med formparameter n og skalaparameter lik $\lambda/(2\lambda) = 1/2$, dvs. denne variable er χ^2-fordelt med $2n$ frihetsgrader.

Oppgave 6

La $T_1, T_2 \ldots, T_{10}$ være 10 uavhengige eksponensielt fordelte variable med parameter $\lambda = 2.5$. Finn 5%- og 95%-kvantilen i fordelingen til middelverdien $\bar{T}$.

Hint: Uttrykk $\bar{T}$ på formen aZ, der a er en kjent konstant og Z er χ^2-fordelt. Sett så opp ulikheten $P(Z < z_\alpha) = 0.05$ (hvor z_α kan finnes i tabellene) og sett inn Z uttrykt ved $\bar{T}$.