Eksponensielle klasser

STK3100 - 1. september 2008

Sven Ove Samuelsen

Plan for 2. forelesning:

1. Definisjoner av eksponensielle klasser
2. Eksempler
3. Forventningsstruktur eksponensielle klasser
4. Likelihood GLM

Definisjon av GLM

Uavhengige responser: $Y_1, Y_2, \ldots, Y_n$
Vektorer av forklaringsvariable $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$
der $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ er p-dimensjonale.

En GLM = Generalisert Lineær Modell er definert ved

- $Y_1, Y_2, \ldots, Y_n$ kommer fra samme eksponensiell klasse (Eksponensielle klasser defineres neste slide, men normalfordelinger, binomiske, Poisson-fordelinger er i eksp. klasser)
- Lineære komponenter (prediktorer) $\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$
- Linkfunksjon $g()$: Med $\mu_i = \mathrm{E}[Y_i]$ kobles forventningen til lineær komponent ved at $g(\mu_i) = \eta_i$

Eksponensielle klasser, de Jong & Heller, Kap. 3

En stokastisk variabel Y sies å ha fordeling i den eksponensielle fordelingsklasse dersom tettheten (pkt.sannsh.) til Y kan skrives på formen

$$f(y;\theta,\phi) = c(y,\phi)\exp\left(\frac{y\theta - a(\theta)}{\phi}\right)$$

der

- θ kalles den kanonisk parameteren
- ϕ kalles spredningsparameteren
- funksjonene $a(\theta)$ og $c(y,\phi)$ er spesifikke for hver enkelt fordeling

Vi kan skrive opp normalfordelingen, binomisk fordeling, poissonfordeling, gammafordeling (eksponensialfordeling) og diverse andre fordelinger på denne formen.

Eksponensielle klasser uten spredningsledd

Noen fordelinger er uten spredningsledd, dvs. $\phi = 1$.
Isåfall blir den eksponensielle klassen gitt ved

$$f(y;\theta) = c(y)\exp(y\theta - a(\theta))$$

Dette gjelder f.eks.

- Poissonfordeling
- fordeling for binære responser, $Y = 1$ eller 0, med $\pi = \mathrm{P}(Y = 1)$
- Binomisk fordeling
- normalfordeling med varians $\sigma^2 = 1$

Eks. Poissonfordeling, $Y \sim$ Po(λ)

Pkt.sannsynlighetene til Y er da

$$f(y;\lambda) = \frac{\lambda^y}{y!}\exp(-\lambda) = \exp(y\log(\lambda) - \lambda - \log(y!)),$$

altså som tetthet i den eksponensielle fordelingsklasse med

- $\theta = \log(\lambda)$
- $a(\theta) = \lambda = \exp(\theta)$
- $c(y) = \frac{1}{y!}$

Eksempel: $Y \sim$ Bin(n, π)

Da har Y har tetthet $f(y;p) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$
som kan omformes til $c(y)\exp(y\theta - a(\theta))$ med

- $\theta = \log(\frac{\pi}{1-\pi})$
- $a(\theta) = n\log(1+\exp(\theta))$
- $c(y) = \binom{n}{y}$

Legg spesielt merke til

- $a'(\theta) = n\frac{\exp(\theta)}{1+\exp(\theta)} = n\pi =$ E$[Y]$
- $a''(\theta) = n\frac{\exp(\theta)}{(1+\exp(\theta))^2} = n\pi(1-\pi) =$ Var$[Y]$

Vi skal se at dette er generelle regler for eksponensielle klasser.

Eksempel: Binær respons

$$Y = \begin{cases} 1 & \text{med sannsynlighet } \pi \\ 0 & \text{med sannsynlighet } 1-\pi \end{cases}$$

Dvs. Y har tetthet

$$f(y;\pi) = \pi^y(1-\pi)^{1-y} = \exp(y\log(\frac{\pi}{1-\pi}) + \log(1-\pi))$$

som kan omformes til $c(y)\exp(y\theta - a(\theta))$ med

- $\theta = \log(\frac{\pi}{1-\pi})$
- som gir $\pi = \pi(\theta) = \frac{\exp(\theta)}{1+\exp(\theta)}$
- som leder til $a(\theta) = -\log(1-\pi(\theta)) = \log(1+\exp(\theta))$
- (dessuten blir $c(y) = 1$)

Eksempel: $Y \sim$ N$(\mu, 1)$

altså normalfordelt med varians $\sigma^2 = 1$. Da blir tettheten

$$\begin{aligned} f(y;\mu) &= \tfrac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}(y-\mu)^2) \\ &= \tfrac{1}{\sqrt{2\pi}}\exp(y\mu - \tfrac{\mu^2}{2} - \tfrac{y^2}{2}) \end{aligned}$$

som kan omformes til $c(y)\exp(y\theta - a(\theta))$ med

- $\theta = \mu$
- $a(\theta) = \frac{\theta^2}{2}$
- $c(y) = \frac{\exp(-\frac{y^2}{2})}{\sqrt{2\pi}}$

Igjen genereres forventning og varians fra $c(\theta)$:

- $a'(\theta) = \theta = \mu =$ E$[Y]$
- $a''(\theta) = 1 =$ Var$[Y]$

Eksponensiell klasse med spredningsparameter

Normalfordelinger med varians $\sigma^2 \neq 1$ kan også settes opp som en eksponensiell klasse, men den naturlige parameteren θ vil avhenge av både μ og σ^2 (vi utelater detaljene).

Vi skal istedet se på slike fordelinger som en eksponensiell klasse med spredningsparameter:

$$f(y;\theta,\phi) = c(y,\phi)\exp(\frac{y\theta - a(\theta)}{\phi})$$

der parameteren ϕ kalles *spredningsleddet* eller spredningsparameteren.

Eksempel: $Y \sim$ $\mathbf{N}(\mu,\sigma^2)$

altså normalfordelt med generell varians σ^2. Da blir tettheten

$$\begin{aligned} f(y;\mu) &= \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}(y-\mu)^2) \\ &= \exp(\frac{y\mu - \mu^2/2 - y^2/2}{\sigma^2} - \log(\sqrt{2\pi}\sigma)) \end{aligned}$$

som kan omformes til $c(y,\phi)\exp((y\theta - a(\theta))/\phi)$ med

- $\theta = \mu$ og $a(\theta) = \frac{\theta^2}{2}$
- og med spredningsledd $\phi = \sigma^2$

Dessuten blir $c(y,\phi) = \frac{\exp(-\frac{y^2}{2\phi})}{\sqrt{2\pi\phi}}$

Kan legge merke til at

- $\mathrm{E}[Y] = \mu = \theta = a'(\theta)$
- $\mathrm{Var}[Y] = \sigma^2 = \phi = \phi a''(\theta)$

Binomisk fordeling med (kjent) spredningsledd

Med $Y \sim \mathrm{Bin}(n,\pi)$ inngår antall forsøk n f.eks. i $a(\theta) = n\log(1+\exp(\theta))$ (der $\theta = \log(\pi/(1-\pi))$.)

Vi kan omforme dette ved å se på *andel* suksesser $Y_0 = \frac{Y}{n}$. Da har Y_0 tetthet $\binom{n}{ny_0}\pi^{ny_0}(1-\pi)^{n(1-y_0)}$ som kan skrives som

$$c(y_0,\phi)\exp(\frac{y_0\theta - a_0(\theta)}{\phi})$$

der $c(y,\phi) = \binom{n}{y}$,

- $a_0(\theta) = \log(1+\exp(\theta))$
- $\phi = \frac{1}{n}$

I dette tilfellet er $\phi = \frac{1}{n}$ kjent, vanligvis er det en parameter som må estimeres.

Momentgenererende funksjon

Hvis Y har tetthet $f(y)$ defineres momentgenererende funksjon ved

$$M_Y(t) = \mathrm{E}[\exp(Yt)] = \int \exp(yt)f(y)dy$$

såsant dette integralet eksisterer for alle t i en omegn om 0.

Vi har da at (STK1100)

- $\mathrm{E}[Y] = M_Y'(0)$
- $\mathrm{E}[Y^2] = M_Y''(0)$
- $\mathrm{E}[Y^r] = M_Y^{(r)}(0)$

samt at $M_Y(0) = 1$.

Momentgenererende funksjon for eksponensiell klasse

Med tetthet $f(y;\theta) = c(y)\exp(y\theta - a(\theta))$ (dvs. eksp. klasse uten spredningsledd) blir momentgenerende funksjon

$$\begin{aligned} M_Y(t) &= \int \exp(yt) f(y;\theta)dy \\ &= \int c(y)\exp(y(t+\theta) - a(\theta))dy \\ &= \exp(a(t+\theta) - a(\theta)) \end{aligned}$$

(fra $\int c(y)\exp(y(t+\theta) - a(t+\theta))dy = \int f(y; t+\theta)dy = 1$
altså integral mhp. en tetthet),
og siden det kan vises at $\int \exp(yt)f(y;\theta)dy < \infty$ for t i en omegn om 0.

Foventning og varians i eksponensiell klass

Vi finner dermed at $M_Y(t) = \exp(a(t+\theta) - a(\theta))$ har deriverte (mhp t)

- $M_Y'(t) = a'(t+\theta)M_Y(t)$
- $M_Y''(t) = a''(t+\theta)M_Y(t) + a'(t+\theta)^2 M_Y(t)$

som gir

- $\text{E}[Y] = M_Y'(0) = a'(\theta)M_Y(0) = a'(\theta)$
- $\text{E}[Y^2] = M_Y''(0) = a''(\theta) + a'(\theta)^2$
- og dermed $\text{Var}[Y] = \text{E}[Y^2] - (\text{E}[Y])^2 = a''(\theta)$

Dette var nettopp hva vi observerte for binomisk fordeling og normalfordeling med $\sigma^2 = 1$.

Eksempel Poissonfordeling

Med $Y \sim$ Po(λ) var kanonisk parameter $\theta = \log(\lambda)$ og $a(\theta) = \exp(\theta)$.
Vi får altså

- $\text{E}[Y] = a'(\theta) = \exp(\theta) = \lambda$
- $\text{var}[Y] = a''(\theta) = \exp(\theta) = \lambda$

(som sikkert er velkjent)

Likelihood-egenskaper

Anta at vi observerer en Y fra $f(y;\theta) = c(y)\exp(y\theta - a(\theta))$.
Da fås

- Log-likelihood-bidrag
  $l(\theta) = \log(f(Y;\theta)) = Y\theta - a(\theta) + \log(c(Y))$
- Score-bidrag $U(\theta) = \frac{\partial l(\theta)}{\partial \theta} = Y - a'(\theta)$
- Bidrag til "informasjon": $\mathcal{J}(\theta) = -\frac{\partial^2 l(\theta)}{\partial \theta^2} = a''(\theta)$

Dermed blir også, ved forventnings og variansreglene for eksp.klasser,

- $\text{E}[U(\theta)] = \text{E}[Y - a'(\theta)] = 0$
- $\text{Var}[U(\theta)] = \text{Var}[Y - a'(\theta)] = a''(\theta) = \mathcal{J}(\theta) = \text{E}[\mathcal{J}(\theta)]$

Dette er generelle likelihood egenskaper, som vi nå har vist for eksp.klasser.

$M_Y(t)$ for eksp. klasse MED spredningsparameter

Tetthet: $f(y;\theta,\phi) = c(y,\phi)\exp(\frac{y\theta - a(\theta)}{\phi})$

Da blir momentgenerende funksjon

$$M_Y(t) = \exp(\frac{a(t\phi+\theta) - a(\theta)}{\phi})$$

Dermed fås

- $M'_Y(t) = a'(t\phi+\theta)M_Y(t)$
- $M''_Y(t) = \phi a''(t\phi+\theta)M_Y(t) + a'(t\phi+\theta)^2 M_Y(t)$

som leder til

- $\text{E}[Y] = M'_Y(t) = a'(\theta)$
- $\text{E}[Y^2] = M''_Y(0) = \phi a''(\theta) + a'(\theta)^2$
- og dermed $\text{Var}[Y] = \phi a''(\theta)$

Eks: Normalfordeling $\mathbf{N}(\mu,\sigma^2)$

Vi fant da at

- $\theta = \mu$ og $a(\theta) = \frac{\theta^2}{2}$
- $\phi = \sigma^2$

som gir kjente formler

- $\text{E}[Y] = a'(\theta) = \theta = \mu$
- $\text{Var}[Y] = \phi a''(\theta) = \phi = \sigma^2$

Eks. Binomisk fordeling $Y \sim$ Bin(n,π)

For $Y_0 = \frac{Y}{n}$ spesifiserte vi

- $\theta = \log(\pi/(1-\pi)$
- $a(\theta) = \log(1+\exp(\theta))$
- $\phi = \frac{1}{n}$

som gir

- $\text{E}[Y_0] = a'(\theta) = \frac{\exp(\theta)}{1+\exp(\theta)} = \pi$
- $\text{Var}[Y_0] = \phi a''(\theta) = \phi\frac{\exp(\theta)}{(1+\exp(\theta))^2} = \frac{1}{n}\pi(1-\pi)$

Likelihood-egenskaper

Med Y fra $f(y;\theta) = c(y,\phi)\exp(\frac{y\theta - a(\theta)}{\phi})$ fås

- $l(\theta) = \log(f(Y;\theta)) = \frac{Y\theta - a(\theta)}{\phi} + \log(c(Y,\phi))$
- $U(\theta) = \frac{\partial l(\theta)}{\partial\theta} = \frac{Y - a'(\theta)}{\phi}$
- $\mathcal{J}(\theta) = -\frac{\partial^2 l(\theta)}{\partial\theta^2} = \frac{a''(\theta)}{\phi}$

som leder til de generelle likelihood-egenskapene

- $\text{E}[U(\theta)] = \frac{1}{\phi}\text{E}[Y - a'(\theta)] = 0$
- $\text{Var}[U(\theta)] = \frac{\phi a''(\theta)}{\phi^2} = \text{E}[\mathcal{J}(\theta)]$

Vi skal senere se utledning av disse formlene for generelle (regulære) likelihooder.

de J & H's bevis for E$[Y] = a'(\theta)$ og Var$(Y) = \phi a''(\theta)$

Vi har

$$\int f'(y;\theta,\phi)dy = \frac{\partial}{\partial\theta}\int f(y;\theta,\phi)dy = \frac{\partial}{\partial\theta}(1) = 0$$

Siden $f'(y;\theta,\phi) = \frac{y-a'(\theta)}{\phi} f(y;\theta,\phi)$ blir dermed

$$0 = \int f'(y;\theta,\phi)dy = \frac{\mathrm{E}[Y] - a'(\theta)}{\phi},$$

dvs. E$[Y] = a'(\theta)$

Tilsvarende fås $0 = \int f''(y;\theta,\phi)dy$ samt

$f''(y;\theta,\phi) = \left[\left(\frac{y-a'(\theta)}{\phi}\right)^2 - \frac{a''(\theta)}{\phi}\right] f(y;\theta,\phi)$ som gir

$$0 = \frac{\mathrm{Var}(Y)}{\phi^2} - \frac{a''(\theta)}{\phi} \Leftrightarrow \mathrm{Var}(Y) = \phi a''(\theta)$$

Variansfunksjon $V(\mu)$

For eksponensielle klasser med spredningsparameter har vi vist

$$\mathrm{Var}(Y) = \phi a''(\theta)$$

Det er en 1-1 sammenheng mellom $\mu =$ E$[Y] = a'(\theta)$ og θ.

Derfor kan vi uttrykke $\theta = \theta(\mu)$ som en funksjon av μ.

Dermed kan vi definere variansfunksjonen

$$V(\mu) = a''(\theta(\mu))$$

slik at Var$(Y) = \phi V(\mu)$.

For de vanligste fordelingene gis uttrykket for $V(\mu)$ direkte.

Variansfunksjon for noen fordelinger

For normalfordelingen er $a''(\theta) = 1$. Dermed blir variansfunksjonen

$$V(\mu) = 1 \qquad \text{(konstantfunksjonen)}$$

For Poissonfordelingen er $a''(\theta) = \exp(\theta) = \mu$, dvs.

$$V(\mu) = \mu \qquad \text{(identitetsfunksjonen)}$$

For binomisk fordeling, med $Y_0 = Y/n$, er $a''(\theta) = \frac{e^\theta}{(1+e^\theta)^2}$ $= \pi(1-\pi)$, dermed blir variansfunksjonen

$$V(\pi) = \pi(1-\pi)$$

Andre medlemmer i den eksponensielle fordelingsklasse

I tillegg til normalfordelinger, binomisk fordelinger og Poisson-fordelinger inneholder den eksponensiell fordelingsklasse bl.a.

- Gamma-fordelinger (og spesielt eksponensialfordelinger)
- Negativt binomisk fordeling (og spesielt geometrisk fordeling)
- Invers gaussisk fordeling
- Trunkerte fordelinger fra andre fordelinger i eksponensiell klasse

Eksponensial- og gammafordeling

Eksponensialfordeling $g(y;\lambda) = \lambda \exp(-\lambda y)$ for $y > 0$
$\mu = \mathrm{E}[Y] = \frac{1}{\lambda}$, $\mathrm{Var}[Y] = \frac{1}{\lambda^2} = \mu^2$
Eksp. klasse uten spredningsledd og med var.fu. $V(\mu) = \mu^2$

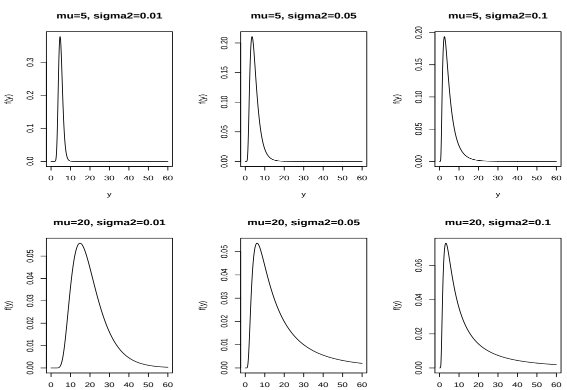
Gammafordeling $g(y;\lambda,\nu) = \frac{y^{\nu-1}\lambda^{\nu}}{\Gamma(\nu)} \exp(-\lambda y)$ for $y > 0$
$\mu = \mathrm{E}[Y] = \frac{\nu}{\lambda}$, $\mathrm{Var}[Y] = \frac{\nu}{\lambda^2} = \frac{\mu^2}{\nu}$
Eksp. klasse med spredningsledd $\phi = 1/\nu$ og var.fu. $V(\mu) = \mu^2$
Hensiktsmessig å reparametrisere til tetthet

$$f(y;\mu,\nu) = \frac{y^{\nu-1}}{\mu^{\nu}\Gamma(\nu)} \exp(-\nu y/\mu)$$

Invers gaussisk fordeling: Y har tetthet, for $y > 0$,

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi y^3 \sigma^2}} \exp(-\frac{1}{2y}\frac{(y-\mu)^2}{\mu^2\sigma^2})$$

der $\mu = \mathrm{E}[Y]$ og $\mathrm{Var}(Y) = \sigma^2\mu^3$. Dette blir en eksponensiell klasse med spredningsledd $\phi = \sigma^2$ og var.fu. $V(\mu) = \mu^3$.



Geometrisk og negativ binomisk fordeling

Geometrisk fordeling: $\mathrm{P}(Y = y) = \pi(1-\pi)^y$ for $y = 0, 1, 2, \ldots$
Eksp. kl. med $\mu = (1-\pi)/\pi$, ingen spredningsparameter og $\mathrm{Var}(Y) = V(\mu) = \mu(1+\mu)$

Negativ binomisk fordeling, $r =$ antall suksesser,
$Y =$ antall Bernoulli forsøk $-r$:

$$\mathrm{P}(Y = y) = \binom{y+r-1}{r-1}\pi^r(1-\pi)^y \qquad \text{for} y = 0, 1, 2, \ldots$$

Fordeling på $y = 0, 1, 2, \ldots$ istedefor $r, r+1, \ldots$ som i STK1100 (Rice)
Dette blir også en eksponensiell klasse med $\mu = \mathrm{E}[Y] = r(1-\pi)/\pi$, uten spredningsparameter, og med $V(\mu) = \mu(1+\mu/r)$.

Negativ binomisk fordeling, generalisering

Vi kan skrive

$$\binom{y+r-1}{r-1} = \frac{\Gamma(y+r)}{y!\Gamma(r)}$$

Dette gir oss muligheten til å generalisere Neg.bin. fordeling til ikke-heltallige r ved punktsannsynligheter

$$\mathrm{P}(Y = y) = \frac{\Gamma(y+r)}{y!\Gamma(r)}\pi^r(1-\pi)^y \qquad \text{for} y = 0, 1, 2, \ldots$$

Med $\kappa = 1/r$ blir dette også en eksp. kl. med $\mu = \mathrm{E}[Y] = r(1-\pi)/\pi$, uten spredningsparameter, og med $V(\mu) = \mu(1+\kappa\mu)$.

Denne fordelingen kan oppstå ved at $Y|\lambda \sim \mathrm{Po}(\lambda)$ der λ er en stokastisk variabel med gammafordeling med forventing μ og formparameter $r = 1/\kappa$.

Oversikt (noen) fordelinger i eksp. klasse

Distrib.	θ	$a(\theta)$	ϕ	E$[Y]$	$V(\mu)$
Bin(n,π)	$\log(\frac{\pi}{1-\pi})$	$\log(1+e^{\theta})$	1	$n\pi$	$n\pi(1-\pi)$
Po(μ)	$\log(\mu)$	$\exp(\theta)$	1	μ	μ
N(μ,σ^2)	μ	$\frac{\theta^2}{2}$	σ^2	μ	1
Gamma(μ,ν)	$-\frac{1}{\mu}$	$-\log(-\mu)$	$\frac{1}{\nu}$	μ	μ^2
Inv.G.(μ,σ^2)	$-\frac{1}{2\mu^2}$	$-\sqrt{-2\theta}$	σ^2	μ	μ^3
NegBin(μ,κ)	$\log(\frac{\kappa\mu}{1+\kappa\mu})$	$\frac{-1}{\kappa}\log(1-\kappa e^{\theta})$	1	μ	$\mu(1+\kappa\mu)$

Estimering i eksponensielle klasser

Anta at $Y_1,\cdots,Y_n$ er uavhengige og har samme tetthet $f(y;\theta,\phi)=c(y,\phi)\exp(\frac{y\theta-a(\theta)}{\phi})$.

Dette gir likelihood $L(\theta)=\prod_{i=1}^n f(Y_i;\theta,\phi)$ og log-likelihood

$$l(\theta)=\sum_{i=1}^n\Big[\frac{Y_i\theta-a(\theta)}{\phi}+\log(c(Y_i,\phi))\Big]$$

samt score-funksjon

$$U(\theta)=\frac{\partial l(\theta)}{\partial\theta}=\frac{1}{\phi}\sum_{i=1}^n[Y_i-a'(\theta)],$$

dvs. maximum likelihood estimatoren (MLE) for θ gis ved å løse ligningen

$$\frac{1}{n}\sum_{i=1}^n Y_i=a'(\hat{\theta})=\hat{\mu}$$

Estimering av spredningsparameter ϕ

kan nå gjøres ved momentprinsippet

$$\hat{\phi}=\frac{1}{n}\sum_{i=1}^n\frac{(Y_i-\hat{\mu})^2}{V(\hat{\mu})}$$

siden

$$\mathrm{E}\Big[\frac{(Y_i-\mu)^2}{V(\mu)}\Big]=\frac{\mathrm{Var}(Y_i)}{V(\mu)}=\phi$$

For normalfordeling blir $\hat{\phi}$ også MLE, men for gammafordeling er MLE for ϕ mer kompleks.

Eksponensiell klasse definert noe annerledes

Y har tetthet/pkt.sanns.

$$f(y;\theta)=\exp(a(y)b(\theta)+c(\theta)+d(y))$$

for gitte funksjoner $a(y),b(\theta),c(\theta),d(y)$.
Ser at med $a(y)=y$ og $b(\theta)=\theta$ fås eksp. kl. uten spredningsparameter.

Ekspempler:

- Paretofordeling $f(y;\theta)=\frac{\theta}{y^{\theta+1}}$ for $y>1$
- Weibullfordeling med kjent formparameter α med tetthet $f(y;\lambda)=\alpha\lambda^{\alpha}y^{\alpha-1}\exp(-(\lambda y)^{\alpha})$ for $y>0$

Kan vise at for $V=a(Y)$ og $\gamma=b(\theta)$ har V tetthet i "vanlig" eksponensiell klasse uten spredningsledd.