STK1100 Obligatorisk oppgave 2

Jonas Thoen Faber

April 8, 2020

Oppgave 1

a)

Den kumulative sannsynlighetsfordelingen for en kontinuerlig tilfeldig variabel X er definert som:

$$F_X(x) = f_X(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

Vi vet at dersom $X \leq \kappa$ så er punktsannsynligheten $f_X(x) = 0$. Så i dette tilfellet er integralet i området $y \in (\kappa, x)$. Vi får da videre:

$$F_X(x) = \theta\kappa^\theta \int_{\kappa}^{x} y^{-\theta-1}dy$$

$$= -\theta\kappa^\theta \frac{1}{\theta}\left[y^{-\theta}\right]_{\kappa}^{x}$$

$$= -\kappa^\theta(x^{-\theta} - \kappa^{-\theta})$$

$$= \kappa^{\theta-\theta} - \kappa^\theta x^{-\theta}$$

$$= 1 - \kappa^\theta x^{-\theta}$$

Medianen $\tilde{\mu}$ er definert som der hvor den kumulative funksjonen når 0.5, det vil si:

$$F_X(\tilde{\mu}) = 1 - \kappa^\theta \tilde{\mu}^{-\theta} = \frac{1}{2}$$

$$\Rightarrow \frac{1}{\tilde{\mu}^\theta} = \frac{1}{2\kappa^\theta}$$

$$\Rightarrow \tilde{\mu} = \sqrt[\theta]{2}\kappa$$

b)

Forventingsverdien for en kontinuerlig stokastisk variabel er gitt som:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Siden sannsynlighetsfunksjonen er lik 0 dersom $x \leq \kappa$ holder det med å integrere fra $\kappa \to \infty$:

$$= \theta\kappa^{\theta} \int_{\kappa}^{\infty} x^{-\theta} dx$$

$$= \theta\kappa^{\theta} \left[\frac{x^{-\theta+1}}{1-\theta} \right]_{\kappa}^{\infty}$$

$$= \theta\kappa^{\theta} \left(\left[\frac{x^{-\theta+1}}{1-\theta} \right]_{x=\infty} - \left[\frac{x^{-\theta+1}}{1-\theta} \right]_{x=\kappa} \right)$$

Siden $\theta > 2$ ser vi at det første leddet vil gå mot null når vi setter inn $x = \infty$. Vi står da igjen med:

$$= \theta\kappa^{\theta} \left(\frac{\kappa^{-\theta+1}}{\theta - 1} \right)$$

$$= \frac{\theta\kappa}{\theta - 1}$$

$$= \frac{\kappa}{1 - 1/\theta}$$

c)

Antar at $\kappa = 400\ 000$ og $\theta = 3$ og finner medianen ved å bruke formelen vi fant i a):

$$\tilde{\mu} = \sqrt[3]{2} \cdot 400000 = 503968$$

Forventet årsinntekt finner vi ved å bruke uttrykket vi fant i b) med samme verdier:

$$E(X) = \frac{400000}{1 - 1/3} = 600000$$

Når vi bruker median luker vi bort de få som tjener veldig mye. I dette tilfelle starter vi fra en årlig lønn på 400 000 så vi luker ikke bort de som tjener veldig lite. Altså vil meridianen være feil-forskjøvet selv om meridianen regner ut en del mindre i lønn. Forventningsverdien tar riktignok med folkegruppen som tjener mye penger men inkluderer også sannsynligheten for at man tjener mye. Altså vil jeg påstå at forventningsverdien gir et best uttrykk for en typsik årsinntekt.

d)

Variansen til en kontinuerlig stokastisk variabel X med sannsynlighetstetthet $f_X(x)$ og forventningsveri $\mu = E(X)$ er gitt ved:

$$\sigma_X^2 = V(X) = E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$$

Vi kan her også integrere fra κ siden $f(x) = 0$ for $x \leq \kappa$. Vi får:

$$= \theta\kappa^\theta \int_\kappa^\infty (x-\mu)^2 \cdot x^{-\theta-1} dx$$

$$= \theta\kappa^\theta \int_\kappa^\infty \left(x^{-\theta+1} - 2\mu x^{-\theta} + \mu^2 x^{-\theta-1}\right) dx$$

$$= -\theta\kappa^\theta \left[\frac{x^{-\theta+2}}{\theta-2} - \frac{2\mu x^{-\theta+1}}{\theta-1} + \frac{\mu^2 x^{-\theta}}{\theta}\right]_\kappa^\infty$$

siden $\theta > 2$ vil alle uttrykkene gå mot 0 når vi setter inn $x = \infty$. Vi står da igjen med:

$$= \theta\kappa^\theta \left(\frac{\kappa^{-\theta+2}}{\theta-2} - \frac{2\mu\kappa^{-\theta+1}}{\theta-1} + \frac{\mu^2\kappa^{-\theta}}{\theta}\right)$$

$$= \theta \left(\frac{\kappa^2}{\theta-2} - \frac{2\mu\kappa}{\theta-1} + \frac{\mu^2}{\theta}\right)$$

$$= \frac{\kappa^2}{1-2/\theta} - \frac{2\mu\kappa}{1-1/\theta} + \mu^2$$

Standardavviket er gitt ved:

$$\sigma_X = \sqrt{V(X)} = \sqrt{\frac{\kappa^2}{1-2/\theta} - \frac{2\mu\kappa}{1-1/\theta} + \mu^2}$$

e)

Nå har vi at $Y = \theta \ln(X/\kappa)$ og vi kjenner sannsynlighetstettheten for X. Vi kan bruke denne til å finne sannsynlighetstettheten til Y ved følgende:

$$f_Y(x) = \theta \ln(\frac{f_X(x)}{\kappa})$$

Vi har definert sannsynlighetstettheten for X for $x > \kappa$ og $x \leq \kappa$. Dersom $x \leq \kappa$ så er $f_X(x) = 0$ som tilsier at $f_Y(x) = udefinert$. Vi får da for $x > \kappa$:

$$f_Y(x) = \theta \ln(\frac{\theta\kappa^\theta x^{-\theta-1}}{\kappa})$$

$$= \theta \ln(\theta\kappa^{\theta-1} x^{-\theta-1})$$

Oppgave 2

a)

Vi har to kontinuerlige stokastiske variable X og Y med en simultan sansynlighetstetthet $f(x,y)$. Med de grensene vi har fått oppgit har vi at:

$$\int_0^1 \int_0^{1-x} f(x,y)dydx = 1$$

$$\Rightarrow \int_0^1 \int_0^{1-x} k(x+2y)dydx = 1$$

$$\Rightarrow k \int_0^1 \left(x(1-x) + (1-x)^2\right) dx = 1$$

$$\Rightarrow k \int_0^1 (1-x)\, dx = 1$$

$$\Rightarrow k \left(1 - \frac{1}{2}\right) = 1$$

$$\Rightarrow k = 2$$

b)

Siden $Y \leq X$ og $x + y \leq 1$ så må $0 \leq y \leq 1/2$ og $y \leq x \leq 1 - y$. Integralet blir da:

$$P(Y \leq X) = \int_0^{1/2} \int_y^{1-y} f(x,y)dxdy$$

$$= 2 \int_0^{1/2} \int_y^{1-y} (x + 2y)\, dxdy$$

$$= 2 \int_0^{1/2} \left[\frac{x^2}{2} + 2yx\right]_y^{1-y} dy$$

$$= 2 \int_0^{1/2} \left(\frac{(1-y)^2}{2} + 2y(1-y) - \frac{y^2}{2} - 2y^2\right) dy$$

$$= \int_0^{1/2} \left(1 - 2y + y^2 + 4y - 4y^2 - y^2 - 4y^2\right) dy$$

$$= \int_0^{1/2} \left(1 + 2y - 8y^2\right) dy$$

$$= \left[y + y^2 - \frac{8}{3} y^3 \right]_0^{1/2}$$

$$= \frac{1}{2} + \frac{1}{4} - \frac{8}{3}\frac{1}{8} = \frac{5}{12}$$

c)

Den marginale sannsynlighetstettheten til X er gitt ved:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy$$

Grensene går i dette tilfelle er $y \in [0, 1-x]$ siden tettheten ellers kun er lik 0:

$$= 2 \int_0^{1-x} (x + 2y) dy$$

$$= 2 \left[xy + y^2 \right]_0^{1-x}$$

$$= 2x(1-x) + 2(1-x)^2$$

$$= 2x - 2x^2 + 2 - 4x + 2x^2 = 2(1-x)$$

Dette gjelder i området $x \in [0,1]$, ellers er den marginale sannsynlighetstettheten alltid 0.

d)

Vi finner sannsynlighetstettheten på samme måte som i c) men nå med hensyn på Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx$$

Her er grensene $x \in [0, 1-y]$ siden $f(x,y)$ kun er definert i dette intervallet slik at vi får:

$$= 2 \int_0^{1-y} (x + 2y) dx$$

$$= 2 \left[\frac{x^2}{2} + 2yx \right]_0^{1-y}$$

$$= (1-y)^2 + 4y(1-y)$$

$$= 1 - 2y + y^2 + 4y - 4y^2 = -3y^2 + 2y + 1$$

e)

Vi vet at to kontinuerlige stokastike variable X og Y er uavhengige dersom $f(x, y) = f_X(x) \cdot f_Y(y)$. Vi har regnet ut de to produktene på høyresiden i oppgavene over og vi får:

$$f_X(x) \cdot f_Y(y) = 2(1-x) \cdot (-3y^2 + 2y + 1)$$

Vi observerer at vi har et ledd der blant annet y^2 og xy er inkludert noe som viser til at:

$$f(x, y) \neq f_X(x) \cdot f_Y(y)$$

Altså er X og Y avhengige kontinuerlige stokastiske variable.

Oppgave 3

a)

Vi har uttrykket for en kumulativ fordeling $F(x)$ som antas er strengt voksende og en uniform stokastisk variabel $U \sim uniform(0, 1)$. Vi setter transformasjonen $X = F^{-1}(U)$ og finner den kumulative fordeligen for X:

$$P(X \leq x) = P(F^{-1}(U) \leq x)$$

$$= P(F(F^{-1}(U)) \leq F(x))$$

$$= P(U \leq F(x))$$

og siden U er uniform, står vi igjen med:

$$= F(x)$$

Altså har $X = F^{-1}(U)$ den kumulative fordelingen $F(x)$.

b)

Fra oppgave 1a) har vi at:

$$F(x) = \begin{cases} 1 - \kappa^\theta x^{-\theta} & x > \kappa \\ 0 & \text{ellers} \end{cases}$$

I forrige deloppgave viste vi at $X = F^{-1}(U)$ har en kumulativ fordeling $F(x)$. Vi bruker dette for å genere observasjoner fra Pareto fordelingen for $x > \kappa$:

$$F(F^{-1}(U)) = F(x)$$

$$\Rightarrow U = \begin{cases} 1 - \kappa^\theta x^{-\theta} & x > \kappa \\ 0 & \text{ellers} \end{cases}$$

$$\Rightarrow x = \frac{\kappa}{\sqrt[\theta]{1-U}}$$

c)

Etter å ha kjørt koden en gang får vi at gjennosnittet $\overline{x} = 600880$ og medianen $\tilde{\mu} = 504570$. Etter å ha kjørt koden flere ganger får vi omtrent de samme verdiene. Verdiene vi får ser vi at stemmer godt overens med medianen og forventningsverdien vi fant i oppgave 1c). Det kan virke som at gjennomsnittet konvergerer mot forventningsverdien som antyder til store talls lov.

d) / e)

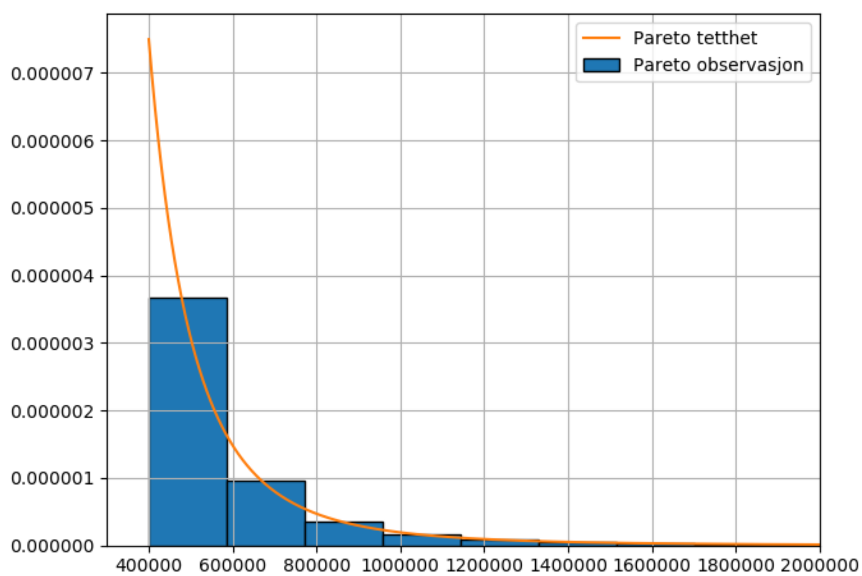

Figure 1: Visualisering av en Pareto fordeling både i form av tettheten og et histogram.

Figure 1 viser et normert histogram av Pareto-fordelte observasjoner som vi fant i forrige deloppgave. Det er også implementert tettheten til Pareto-fordelingen som vi ser samsvarer med histogrammet når vi bruker de samme verdiene for κ og θ. Dette viser til at dersom vi har en stokastisk variabel U som er uniformt fordelt mellom 0 og 1 så kan vi generere observasjonene fra Pareto fordelingen.

Kode som tilhører oppgave 3

```python
import numpy as np, matplotlib.pyplot as plt, random as r


######### oppgave c) #########
# genererer et uniformt tilfeldig tall mellom 0 og 1
def U():
    return r.random()

kap = 4e5
theta = 3

N = int(1e5)
X = np.zeros(N)

# Regner ut verdier for x mellom kappa og 2e6
for i in range(len(X)):
    X[i] = kap / (1-U())**(1/theta)

med = np.median(np.sort(X))
snitt = sum(X)/len(X)
print("Median:          %i" % med)
print("Gjennomsnitt:    %i" % snitt)
print("")

"""
In [24]: run Pareto_fordeling.py
Median:          504570
Gjennomsnitt:    600880

Vi ser at verdiene vi faar er veldig naer de vi fant i oppgave 1c) som tilsvarer:
Median:          503968
Forventet:       600000
"""


######### oppgave d) #########
plt.xlim(int(3e5), int(2e6))
plt.hist(X, bins=200, normed=True, edgecolor="black", label="Pareto observasjon")
#plt.show()



######### oppgave e) #########
def f(x):
    return theta * kap**(theta) * x**(-theta-1)
x = np.linspace(kap,2e6,N)
plt.plot(x,f(x), label="Pareto tetthet")
plt.legend(); plt.grid(); plt.xscale("linear")
plt.show()
```