STK 1100

Obligatorisk oppgave 2

Jonas Semprini Næss

23. april 2020

Oppgave 1:

a.)

La X være årsinntekten til en tilfeldig valgt person i en befolkningsgruppe. Det er vanlig å anta at X er Pareto-fordelt, det vil si at X har sannsynlighetstettheten

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

Vis at den kumulative sannsynlighetsfordelingen til X er gitt ved:

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

Løsning:

Vi husker at den kumulative sannsynlihetsfordelingen for en kontinuerlig stokastisk variabel X er gitt ved.

$$F(x) = \int_{-\infty}^{x} f(u)\ du$$

Da får vi

$$F(x) = \int_{-\infty}^{k} f(u)\ du + \int_{k}^{x} f(u)\ du$$

Integralet $\int_{-\infty}^{k} f(u)\ du$ vil sumere til 0 ettersom at sannsynlighetstettheten til X er null for alle verdier x mindre enn κ. Det gir dermed

$$F(x) = \int_{\kappa}^{x} f(u)\ du$$

$$F(x) = \int_{\kappa}^{x} \theta\kappa^{\theta} u^{-\theta-1}\ du$$

$$F(x) = \theta\kappa^{\theta} \int_{\kappa}^{x} u^{-\theta-1}\ du$$

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

$$F(x) = -\kappa^{\theta} \left[x^{-\theta} - (\kappa^{-\theta})\right]$$

$$F(x) = 1 - \frac{k^{\theta}}{x^{\theta}}$$

Som var det vi skulle vise.

b.)

Finn $E(x)$

Løsning:

Vi husker at forventingen til en kontinuerlig stokastisk variabel X er gitt ved

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

hvor $f(x)$ er sannsynlighetstettheten til X. Bruker vi ressonementet får vi

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

Igjen vil integralet $\int_{-\infty}^{\kappa} xf(x)\ dx$ summere til null gitt at X har en sannsynlighetstetthet som gir utslag for alle $x > \kappa$. Videre får vi

$$E(X) = \int_{\kappa}^{\infty} xf(x)\ dx$$

$$E(X) = \int_{\kappa}^{\infty} x\theta\kappa^{\theta}x^{-\theta-1}\ dx$$

$$E(X) = \theta\kappa^{\theta}\int_{\kappa}^{\infty} xx^{-\theta-1}\ dx$$

$$E(X) = \theta\kappa^{\theta}\int_{\kappa}^{\infty} xx^{-\theta}\frac{1}{x}\ dx$$

$$E(X) = \theta\kappa^{\theta}\int_{\kappa}^{\infty} x^{-\theta}\ dx$$

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

Studerer vi $\frac{1}{\theta-1}x^{-\theta+1}$ når $x \to \infty$ ser vi at utrykket går mot null ettersom at nevneren vokser raskere mot uendelig enn telleren. Det gir oss

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

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

$$E(X) = \underline{\underline{\frac{\kappa\theta}{\theta-1}}}$$

Vi kan dermed skrive $E(x)$ som

$$E(X) = \begin{cases} \frac{\kappa\theta}{\theta-1} & \theta > 1, \\ \infty & \theta \leq 1 \end{cases}$$

c.)

I dette punktet antar vi at $\kappa = 400\ 000$ kroner og $\theta = 3$. Hva blir da median årsinntekt og forventet årsinntekt? Hvilken av de to størrelsene gir etter din mening best utrykk for den "typiske årsinntekten" i befolkningsgruppen? Svaret skal begrunnes.

Løsning:

Vi starter med å beregne forventet årsinntekt. Setter vi inn verdiene for κ og θ

får vi.

$$
\begin{aligned}
E(X) &= \frac{\kappa\theta}{\theta-1} \\
E(X) &= \frac{400\ 000 \cdot 3}{2} \\
E(x) &= \underline{\underline{600\ 000}}
\end{aligned}
$$

Altså er den forventede årsinntekten til en tilfeldig person i befolkningsgruppen 600 000 kroner.
Videre ønsker vi å regne på medianen. Dermed må vi se på 50-persentilen til den kumulative sannsynlighetsfordeling og løse med hensyn på x. Det gir

$$
\begin{aligned}
F_X(x) &= \begin{cases} 1-\kappa^{\theta}x^{-\theta} & x>\kappa, \\ 0 & \text{ellers.} \end{cases} \\
F_X(x) &= 1-\kappa^{\theta}x^{-\theta} = \frac{1}{2} \\
F_X(x) &= 1-400\ 000^3x^{-3} = \frac{1}{2} \\
F_X(x) &= 400\ 000^3 \cdot 2 = x^3 \\
F_X(x) &= \sqrt[3]{400\ 000^3 \cdot 2} = \sqrt[3]{x^3} \\
F_X(x) &= x = \underline{\underline{503968.42}}
\end{aligned}
$$

Vi får dermed at median årsinntekten i befolkningsgruppen er 503968.42 kroner.

For å avgjøre hvilken model som representerer den typiske årsinntekten best er det viktig å se på hvordan den kumulative sannsynlighetsfunksjonen operer. Det er forstått at lønnsforskjellene er moderate ettersom at $\theta = 3$, og alltid er større enn to. Dette betyr at funksjonen øker raskt mot én, langs y-aksen som svarer til sannsynligheten for at så og så mange tjener så og så mye. Dermed vil en stor del av befolkningen tjene mer enn minsteinntekten, hvilket tyder til at forventningsverdien($E(X)$) gjenspeiler den typiske årsinntekten mest presist.

d.)

Finn variansen og standardavviket til X

Løsning:

Husker at variansen for en kontinuerlig stokastisk variabel er gitt ved

$$V(X) = E(X^2) - (E(X))^2$$

Starter med å finne $E(X^2)$

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

$$E(X^2) = \int_{\kappa}^{\infty} x^2 f(x)\ dx$$

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

$$E(X^2) = \theta\kappa^{\theta} \int_{\kappa}^{\infty} x^{-\theta-1+2}\ dx$$

$$E(X^2) = \theta\kappa^{\theta} \int_{\kappa}^{\infty} x^{-\theta+1}\ dx$$

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

Følger samme ressonement som i oppgave b.) og ser på utrykket $\frac{1}{\theta+2}x^{-\theta+2}$ når $x \to \infty$. Da vil vi få et $\frac{\infty}{\infty}$ utrykk som betyr at vi kan teste grensen med L'Hopital. Gjentar vi prosessen to ganger får vi et $\frac{k}{\infty}$ utrykk hvor k simpelthen er en konstant i $\mathbb{R}$. Dermed sumerer utrykket til null. Det gir videre

$$E(X^2) = -\theta\kappa^{\theta} \left(-\frac{\kappa^2}{(\theta-2)\kappa^{\theta}}\right)$$

$$E(X^2) = \underline{\frac{\kappa^2\theta}{\theta-2}}$$

Utover er fremgangsmåten enkel, vi subtraherer $(E(X))^2$ og får

$$V(X) = \frac{\kappa^2\theta}{\theta-2} - (E(X))^2$$

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

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

$$V(X) = \frac{\kappa^2\theta(\theta-1)^2}{(\theta-2)(\theta-1)^2} - \frac{\kappa^2\theta^2(\theta-2)}{(\theta-1)^2(\theta-2)}$$

$$V(X) = \frac{\kappa^2\theta(\theta-1)^2 - \kappa^2\theta^2(\theta-2)}{(\theta-2)(\theta-1)^2}$$

$$V(X) = \frac{\kappa^2\theta}{(\theta-1)^2(\theta-2)}$$

For alle $\theta > 1 \wedge \theta > 2$.
Likledes ønsker vi å finne standaravviket til X og da husker vi at standaravviket er gitt ved

$$SD(X) = \sqrt{V(X)}$$

Hvilket gir

$$SD(X) = \sqrt{\frac{\kappa^2\theta}{(\theta-1)^2(\theta-2)}}$$
$$SD(X) = \frac{\kappa}{\theta-1}\sqrt{\frac{\theta}{\theta-2}}$$

Setter vi inn verdiene fra b.) får vi at variansen blir

$$V(X) = \frac{\kappa^2\theta}{(\theta-1)^2(\theta-2)}$$
$$V(X) = \frac{400\ 000^2 \cdot 3}{(2)^2(1)}$$
$$V(X) = \underline{\underline{1.2 \cdot 10^{11}}} \text{ kroner}$$

Og standaravviket gir

$$SD(X) = \frac{400\ 000}{2}\sqrt{\frac{3}{1}}$$
$$SD(X) = 346410.1615 \approx \underline{\underline{346410.16}} \text{ kroner}$$

e.)

La $Y = \theta \ln(X/\kappa)$. Bestem sannsynlighetstettheten til Y. Hvilken kjent sannsynlighetstetthet er det?

Løsning:

Vi er gitt en transformasjon Y av X som betyr at vi ønsker å finne X utrykt ved Y. Løser vi dermed opp utrykket får vi

$$Y = \theta \ln(X/\kappa)$$
$$e^{\frac{Y}{\theta}} = \frac{X}{\kappa}$$
$$X = e^{\frac{Y}{\theta}}\kappa$$

Videre ønsker vi å finne sannsynlighetstettheten til $f_Y(y)$ så vi plugger inn resultatet i $P(X=x)$ hvor $h(y) = X = e^{\frac{Y}{\theta}}\kappa$. Det gir

$$
\begin{aligned}
f_X(x) &= \theta\kappa^\theta x^{-\theta-1} \\
f_Y(y) &= f_X(h(y)) = \theta\kappa^\theta h(y)^{-\theta-1} \\
f_Y(y) &= \theta\kappa^\theta \left(e^{\frac{y}{\theta}}\kappa\right)^{-\theta-1} \\
f_Y(y) &= \theta\kappa^\theta \left(\kappa^{-\theta-1}e^{\frac{y(-\theta-1)}{\theta}}\right) \\
f_Y(y) &= \frac{\theta}{\kappa}e^{\frac{y(-\theta-1)}{\theta}}
\end{aligned}
$$

Så husker vi at for å finne sannsynlighetstettheten til en vilkårlig transformasjon Y av X har vi utrykket $f_Y(y) = f_X((h(y))\left|h'(y)\right|$ dermed får vi sluttvis

$$
\begin{aligned}
f_Y(y) &= \frac{\theta}{\kappa}e^{\frac{y(-\theta-1)}{\theta}}h'(y) \\
f_Y(y) &= \frac{\theta}{\kappa}e^{\frac{y(-\theta-1)}{\theta}}\frac{\kappa}{\theta}e^{\frac{y}{\theta}} \\
f_Y(y) &= e^{\frac{y(-\theta-1)+y}{\theta}} \\
f_Y(y) &= e^{\frac{y(-\theta-1+1)}{\theta}} \\
f_Y(y) &= \underline{\underline{e^{-y}}}
\end{aligned}
$$

Vi ser derfor at $F_y(y)$ representerer eksponential fordelingen med skalaparameter $\lambda = 1$.

Oppgave 2:

a.)

De stokastiske variablene X og Y har simultan sannsynlighetstetthet

$$
f(x,y) = \begin{cases} k(x+2y) & 0 \le x \le 1,\ 0 \le y \le 1,\ x+y \le 1, \\ 0 & \text{ellers.} \end{cases}
$$

Der k er en konstant.

Vis at $k = 2$.

Løsning:

Vi husker at sannsynligheten for at alle x, y forekommer innenfor definisjonsmengden er gitt ved

$$f(x,y) = \iint_{x,y \in D_f} f(x,y)\ dxdy = 1$$

$$f(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\ dxdy = 1$$

Videre har vi fått oppgitt at x, y er begrenset innenfor intervallet $[0,1]$ og at funksjonen $f(x,y)$ er begrenset av linjen $y = 1 - x$. Dermed blir integralet følgelig

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

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

$$f(x,y) = k \int_0^1 \left[xy + y^2\right]_0^{1-x}\ dx = 1$$

$$f(x,y) = k \int_0^1 \left[x(1-x) + (1-x)^2\right]\ dx = 1$$

$$f(x,y) = k \int_0^1 [1-x]\ dx = 1$$

$$f(x,y) = \left[x - \frac{1}{2}x^2\right]_0^1 = 1$$

$$f(x,y) = k(1 - \frac{1}{2}) = 1$$

$$f(x,y) = k = 2$$

Som var det vi skulle vise.

b.)

Bestem $P(Y \leq X)$

Løsning:

Vi husker at $f(x,y)$ er begrenset innenfor linjen $y = 1 - x$ som betyr at vi får

$$P(Y \leq X) = P(1 - X \leq X)$$

$$P(Y \leq X) = P(\frac{1}{2} \leq X)$$

Hvilket betyr at $Y \in [0, \frac{1}{2}]$.

Videre ønsker vi å finne hvor X er begrenset. Det er åpentbart at X må være begrenset av Y nedenifra ettersom at $X \geq Y$ for alle $X, Y \in D_f$. Ovenifra vil linjen $x = 1 - y$ svare til hvilke X-verdier som ligger innenfor linjen $Y = X$ og dermed vil $x = 1 - y$ være øvre grense. Da får vi følgende integral

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

$$P(Y \leq X) = \int_0^{\frac{1}{2}} \int_y^{1-y} (2x + 4y)\ dxdy$$

$$P(Y \leq X) = \int_0^{\frac{1}{2}} \Big[x^2 + 4yx\Big]_y^{1-y}\ dy$$

$$P(Y \leq X) = \int_0^{\frac{1}{2}} \Big[(1-y)^2 + 4y(1-y) - (y^2 + 4y^2)\Big]\ dy$$

$$P(Y \leq X) = \int_0^{\frac{1}{2}} 1 - 8y^2 + 2y\ dy$$

$$P(Y \leq X) = \left[y - \frac{8}{3}y^3 + y^2\right]_0^{\frac{1}{2}}$$

$$P(Y \leq X) = \frac{1}{2} - \frac{8}{24} + \frac{1}{4} = \frac{10}{24} = \underline{\underline{\frac{5}{12}}}$$

c.)

Vis at den marginale sannsynlighetstettheten til X er

$$f_X(x) = \begin{cases} 2(1-x) & 0 \leq X \leq 1, \\ 0 & \text{ellers.} \end{cases}$$

Husker at den marginale sannsynlighetstetthet til X er gitt ved

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

Vi beregner altså det totale integralet av $f(x, y)$ over alle mulige $y \in D_f$. Videre husker vi at y er begrenset i intervallet $[0, 1 - x]$, som gir oss grenseverdiene til integralet. Da får vi

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

$$f_X(x) = \int_0^{1-x} (2x+4y)\ dy$$

$$f_X(x) = \Big[2xy + 2y^2\Big]_0^{1-x}$$

$$f_X(x) = \Big[2x(1-x) + 2(1-x)^2\Big]$$

$$f_X(x) = \Big[2x - 2x^2 + 2 - 4x + 2x^2\Big]$$

$$f_X(x) = [2-2x]$$

$$f_X(x) = 2(1-x)$$

Som gjelder for alle $X \in [0,1]$.

d.)

Bestem den marginale sannsynlighetstettheten til Y

Løsning:

Vi bruker samme fremgangsmåte som i c.):

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

Bruker videre at Y er begrenset over intervallet $[0, 1-y]$ ettersom at vi integrerer over alle $X \in D_f$ hvilket er inneholdt i linjen $x = 1-y$. Det gir

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

$$f_Y(y) = \int_0^{1-y} (2x+4y)\ dx$$

$$f_Y(y) = \Big[x^2 + 4xy\Big]_0^{1-y}$$

$$f_Y(y) = \Big[(1-y)^2 + 4y(1-y)\Big]$$

$$f_Y(y) = \Big[-3y^2 + 2y + 1\Big]$$

Som gjelder for alle $Y \in [0,1]$, og vi kan skrive tettheten som

$$f_Y(y) = \begin{cases} -3y^2 + 2y + 1 & 0 \le Y \le 1, \\ 0 & \text{ellers.} \end{cases}$$

e.)

Er X og Y uavhengige? Svaret skal begunnes!

Løsning:

For at to stokastiske variable skal være uavhengig må de oppfyllet kravet $P(A \cap B) = P(A) \cdot P(B)$. Dette gjelder også for flervariable sannsynlighetstettheter, hvilket gir.

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

Er dessuten variablene avhengig, vil ikke ulikheten

$$f(a \leq X \leq b, c \leq Y \leq d) = f_X(a \leq X \leq b) \cdot f_Y(c \leq Y \leq d)$$

holde.

Følger vi ressonementet får vi følgende

$$f(x,y) = f_X(x,y) \cdot f_Y(x,y)$$
$$f(x,y) = 2(1-x)(-3y^2 + 2y + 1)$$
$$f(x,y) = -6y^2 + 6y^2x + 4y - 4yx - 2x + 2$$

Hvilket gir $f(x,y) \neq f_X(x,y) \cdot f_Y(x,y)$ som betyr at X og Y er avhengig.

Oppgave 3:

a.)

La $F(x)$ være en kumulativ fordeling, og anta at $U \sim \text{uniform}(0,1)$. Vis at da har $X = F^{-1}(U)$ kumulativ fordeling $F(x)$

Løsning:

Bevis:

Gitt at vi har en kumulativ fordelingsfunksjon $X = F(x)$ som er strengt voksende og høyre kontinuerlig (monoton) med en venstre nedre grense, vil den følgelig være injektiv/inverterbar.

La så, ξ definere

$$\inf\{x : \ U \leq F(x)\}$$

hvor $U \in (0,1)$. Dermed vil $\boxed{\xi(u) = F^{-1}(x)}$ for alle $x \in \mathbb{R}$. Videre har vi fra definisjonen

$$F(x) \geq u \iff x \geq \xi(u)$$

Hvilket gir

$$F(x) \geq F(\xi(u))$$

La så $A_u = \{k : F(k) \geq u\}$ betegne mengden hvor $\forall k \in \mathbb{R} : k \geq F^{-1}(u)$. Da må det finnes en delmengde $k_n \subseteq A_u$ slik at $k_n = \inf\{A_u\} = \xi(u)$. Dermed har vi at $F(k_n) = F(\xi(u))$ for alle k, og følgelig er $F(\xi(u)) \geq u$.

Sluttvis husker vi at $F(x) \geq F(\xi(u))$, hvilket betyr $F(x) \geq F(\xi(u)) \geq u$, som er ekvivalent med å si at $F(x) \geq u$ hvis og bare hvis $x \geq u$. Tar vi dermed sannsynligheten $P(X \leq x)$ får vi

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

som beviser påstaden om at $X = F^{-1}(U)$ har kumulativ fordeling $F(x)$.

b.)

Bruk resultatene i punkt a.) i denne oppgaven og punkt a i Oppgave 1 til å angi en fremgangsmåte for å generere observasjoner fra Pareto fordelingen

Løsning:

Vi viste i oppgave a.) at $F^{-1}(U)$ er den inverse funksjonen til X, eller en såkalt "kvantil-funksjon". Benytter man resultatet kan en videre finne det generelle utrykket for kvantil-funksjonen, utrykt ved U.

$$\begin{aligned}
X &= F^{-1}(U) \\
U &= F(X) \\
U &= 1 - \left(\frac{\kappa}{X}\right)^{\theta} \\
\left(\frac{\kappa}{X}\right)^{\theta} &= 1 - U \\
\frac{\kappa}{X} &= (1-U)^{\frac{1}{\theta}} \\
X &= \frac{\kappa}{1-U)^{\frac{1}{\theta}}}
\end{aligned}$$

Denne kvantil-funksjonen kan benyttes for å generere observasjoner svarende til Pareto fordelingen, gitt at man har et kapabelt verktøy som generer tilfeldige tall mellom $(0, 1)$, uniformt fordelt.

c.)

Bruk fremgangsmåten i forrige punkt til å generere 10 000 *observasjoner fra Pareto-fordelingen med* $\kappa = 400\ 000$ *kroner og* $\theta = 3$. *Beregn gjennomsnittet og medianen av de genererte observasjonene, og sammenlign med resultatene i*

Oppgave 1, punkt c.

Løsning:

For å løse denne oppgaven lager vi et python-script ut i fra resultatet i b.) og de gitte verdiene i oppgaveteksten. Scriptet kan se følgende ut

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import statistics

theta = 3
k = 400000
u = np.random.uniform(0,1, 10000)

fxu = (k)/((1 - u)**(1/theta))

median = np.median(fxu)
mean = np.mean(fxu)

print(f"Median: {median:.2f} Gjennomsnitt: {mean:.2f}")
```

Her har vi benyttet oss av de innebygde funksjonene np.median og np.mean for å beregne median og gjennomsnitt, men medianen kan også finnes ved å sette $U = \frac{1}{2}$, ettersom at kvantil-funksjonen generer persentiler i henhold til ulike verdier av U.

En viktig bemerkning å legge til er at verdiene for gjennomsnitt og median vil variere fra gang til gang. Dette har med at maskinen genererer nye verdier mellom 0 og 1 hver gang programmet kjøres. Et vilkårlig forsøk ga dermed verdiene

$$\text{Median} = 503223.6$$
$$\text{Gjennomsnitt} = 596440.49$$

Hvilket ikke er langt i fra verdiene vi fikk for medianen og forventningen til $f_X(x)$ i Oppgave 1 c.). Det betyr dermed at kvantil-funksjonen er god tilnærming til den kumulative fordelingsfunksjonen.

d.)

Lag et normert histogram av de Pareto-fordelte observasjonene fra punkt c.) (dvs. et histogram der arealet av alle stolpene til sammen er lik én.

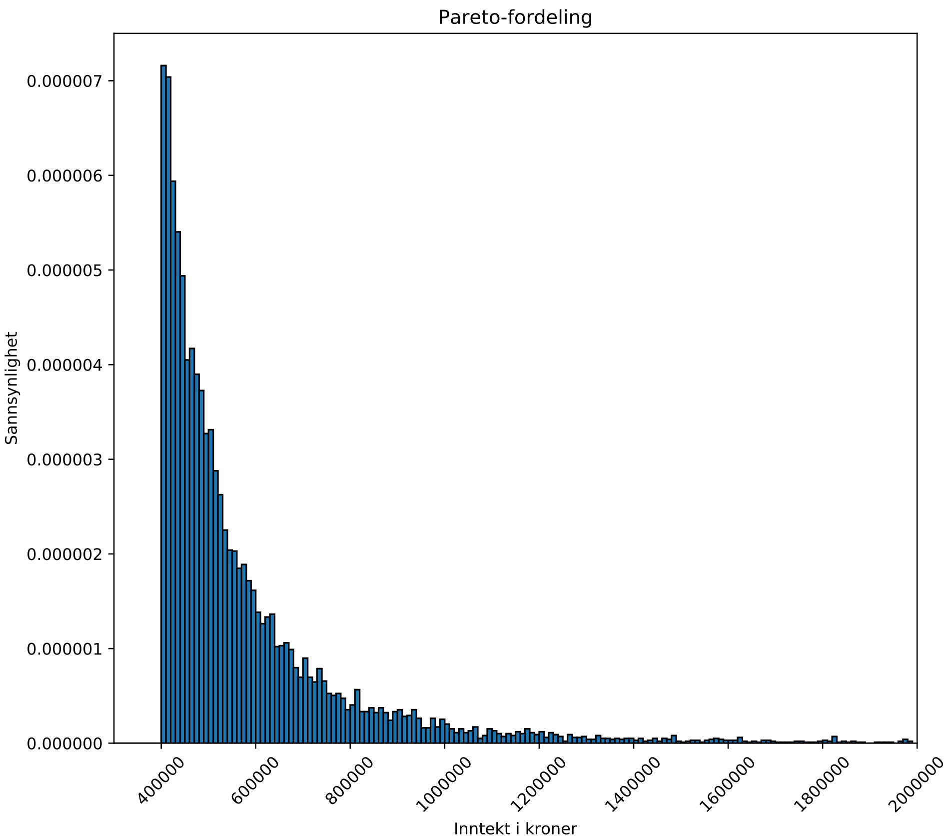
Løsning:

Vi lager et utvidet script fra det i c.) som kan se noe slik ut

```python
import numpy as np
```

```
import matplotlib.pyplot as plt
import pandas as pd
import statistics

theta = 3
k = 400000
u = np.random.uniform(0,1, 10000)

fxu = (k)/((1 - u)**(1/theta))

median = np.median(fxu)
mean = np.mean(fxu)

print(f"Median: {median:.2f} Gjennomsnitt: {mean:.2f}")

fig = plt.figure(figsize=(9,8))
plt.title('Pareto-fordeling')
plt.xlim(300000,2000000)
plt.ylim(0, (7.5*10**-6))
plt.xlabel('Inntekt i kroner')
plt.ylabel('Antall forekomster av gitt inntekt')
plt.xticks(rotation=45)
plt.hist(fxu, bins=range(300000,2000000, 10000), density=True,edgecolor='black')
plt.savefig('pareto.pdf')
plt.show()
```

Kjører vi scriptet får vi følgende plot



Plot av normert histogram for $X = \frac{\kappa}{1-U)^{\frac{1}{\theta}}}$

e.)

Tegn tettheten til Pareto-fordelingen med $\kappa = 400\ 000$ kroner og $\theta = 3$ i samme figur som histogrammet og kommenter resultatet.

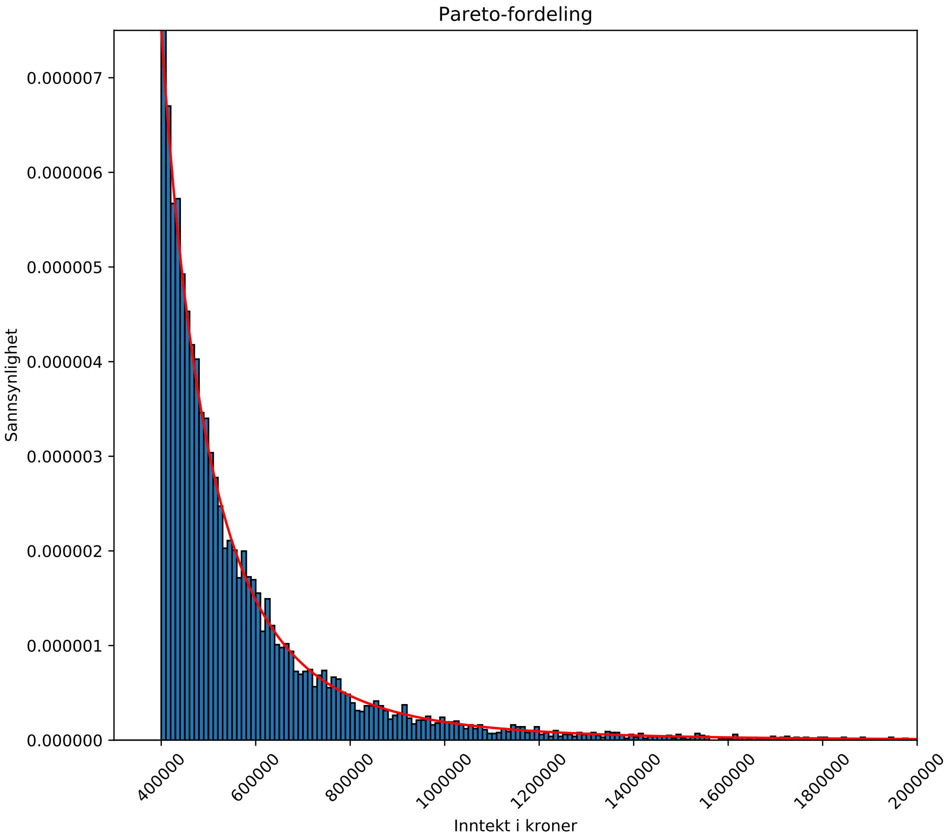
Løsning

Utvider nok en gang foregående script, og implementer sannsynlighetstettheten fra oppgave 1 a.) inn i scriptet

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import statistics

theta = 3
k = 400000
u = np.random.uniform(0,1, 10000)
x = np.linspace(300000, 2000000, 10000)
```

```python

fxu = (k)/((1 - u)**(1/theta))

median = np.median(fxu)
mean = np.mean(fxu)

def fx (x):
    pareto = (theta*k**(theta))/(x**(theta + 1))
    return pareto

print(f"Median: {median:.2f} Gjennomsnitt: {mean:.2f}")

fig = plt.figure(figsize=(9,8))
plt.title('Pareto-fordeling')
plt.xlim(300000,2000000)
plt.ylim(0, (7.5*10**-6))
plt.xlabel('Inntekt i kroner')
plt.ylabel('Sannsynlighet')
plt.xticks(rotation=45)
plt.hist(fxu, bins=range(300000,2000000, 10000), density=True,edgecolor='
    black')
plt.plot(x, fx(x), color='red', label='f_X(x)')
plt.savefig('pareto.pdf')
plt.legend()
plt.show()
```

Ved gjennomkjøring gir scriptet følgende plot



Plot av normert histogram for $X = \frac{\kappa}{1-U)^{\frac{1}{\theta}}}$ og $f_X(x) = \theta\kappa^\theta x^{-\theta-1}$

Som skissert ser vi at histogrammet gjennspeiler sannsynlighetstettheten til Pareto-fordeling, og ved en enda finere oppstykkning av søylene vil histogrammet mer eller mindre glid naturlig over i sannsynlighetstettheten. Dermed er det altså inneforstått at kvantil-funksjonen simulerer sannsynlighetstettheten i store trekk.