# To-dimensjonale kontinuerlige fordelinger

## Noen resultater for diskrete fordelinger

Vi har tidligere definert punktsannsynligheten $p(x, y)$ for en todimensjonal variabel $(X, Y)$ som

$$p(x, y) = P(\{X = x\} \cap \{Y = y\}). \tag{1}$$

Ved å betrakte alle mulige utfall $(x, y)$ som en oppdeling av utfallsrommet har vi

$$\sum_{\text{alle } (x,y)} p(x, y) = 1. \tag{2}$$

Hvis $A$ er et vilkårlig område i $(x, y)$-plane har vi

$$P(\{(X, Y) \in A\}) = \sum_{(x,y)\in A} p(x, y). \tag{3}$$

Marginalfordelingene finnes også ved direkte å anvende loven om total sannsynlighet

$$p_X(x) = P(X = x) = \sum_{\text{alle } y} P(\{X = x\} \cap \{Y = y\}) = \sum_{y} p(x, y) \tag{4}$$

hvor alle verdier av $y$ her representerer en oppdeling av utfallsrommet. Vi kan her bytte om $x$ og $y$

$$p_Y(y) = P(Y = y) = \sum_{\text{alle } x} P(\{X = x\} \cap \{Y = y\}) = \sum_{x} p(x, y). \tag{5}$$

Forventningsverdien til en vilkårlig funksjon $g(X, Y)$ er gitt ved

$$\mathrm{E}g(X,Y) = \sum_{\text{alle } (x,y)} g(x,y)p(x,y). \tag{6}$$

Den betingede fordelingen til $X$ gitt $\{Y = y\}$ er definert ved de betingede sannsynlighetene

$$p_{X|Y}(x|y) = \frac{P(\{X = x\} \cap \{Y = y\})}{P(\{Y = y\})} = \frac{p(x,y)}{p_Y(y)}. \tag{7}$$

Dette er en vanlig endimensjonal fordeling (hvor $y$ inngår som en parameter) som f. eks. har en forventningsverd $\mathrm{E}(X|Y = y)$ og varians $\mathrm{var}(X|Y = y)$ som er funksjoner av verdien $y$ som vi betinger med hensyn på.

Kovariansen mellom $X$ og $Y$ er definert som $\mathrm{cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}X)(Y - \mathrm{E}Y)] = \mathrm{E}(XY) - \mathrm{E}X\mathrm{E}Y$ kan beregnes som

$$\mathrm{cov}(X,Y) = \sum_{\text{alle } (x,y)} xy\ p(x,y) - \mathrm{E}X\mathrm{E}Y. \tag{8}$$

# Definisjon av to-dimensjonale kontinuerlige fordelinger

Resultater som er analoge med likning (1)-(8) gjelder også for kontinuerlige variable. Punktsannsynligheten $p(x, y)$ erstattes da med en sannsynlighetstetthet $f(x, y)$ (multiplisert med $dxdy$) og summer erstattes med integraler.

En to-dimensjonal sannsynlighetstetthet $f(x, y)$ er en funksjon i to variable. Integralet av en slik funksjon over et område $A$ i planet kan tolkes som volumet av den mengden som har $A$ som grunnflate og ligger under den flaten som $f(x, y)$ definerer og uttrykkes matematisk

$$I = \iint_A f(x,y)dxdy. \tag{9}$$

En to-dimensjonal sannsynlighetstetthet $f(x, y) \geq 0$ er definert i analogi med likning (3) ved at

$$P(\{(X, Y)\} \in A) = \iint_A f(x, y)dxdy. \tag{10}$$

Sannsynligheten for at $(X, Y)$ er et punkt i $(x, y)$-planet skal pr. definisjon være lik 1. Dette gir analogien til likning (2)

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1. \tag{11}$$

*Eksempel a, to-dimensjonal rektangulærfordeling*

Anta at alle utfall med $X$ og $Y$ i intervallet [0,1] er like sannsynlige. Vi modellerer dette ved å la den to-dimensjonale sannsynlighetstettheten være konstant over det aktueller området, dvs.

$$f(x, y) = 1 \tag{12}$$

for $0 \leq x \leq 1$ og $0 \leq y \leq 1$ og null utenfor dette området. Sannsynligheten for at $(X, Y) \in A$, der $A$ ligger i det mulige området, blir nå lik arealet til $A$ fordi høyden (funksjonsverdien $f(x, y)$) er lik 1 over hele dette området.

*Eksempel b*

La nå $f(x, y)$ være proporsjonal med $xy$ i området begrenset av linjene $x = 0$, $y = 0$ og $x + y = 1$, og null utenfor denne trekanten.

For en gitt verdi av $X$ lik $x$, kan da $Y$ ta verdier i intervallet $[0, 1 - x]$. Det totale volumet under $f(x, y)$ kan da beregnes som

$$V = \int_0^1 \left[ \int_0^{1-x} f(x, y)dy \right] dx \tag{13}$$

hvor vi i det innerste integralet $\int_0^{1-x} f(x, y)dy$ holder $x$ konstant og lar $y$ være integrasjonsvariabel. For denne enkle modellen blir dette integralet

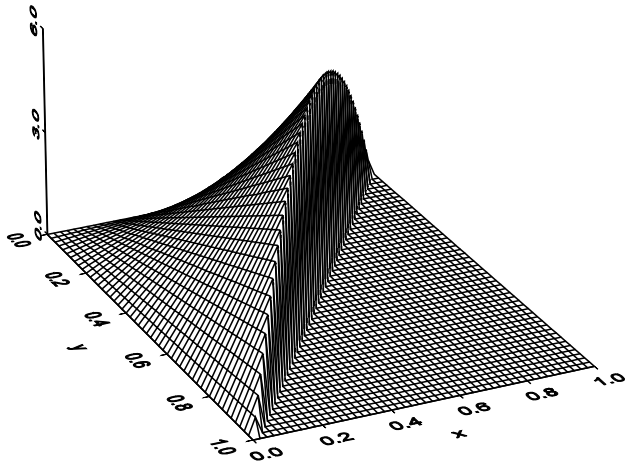
$$\int_0^{1-x} f(x,y)dy = cx\int_0^{1-x} ydy = \frac{1}{2}x(1-x)^2. \tag{14}$$

Ved å sette dette inn i (13) finner vi volumet

$$V = \int_0^1 \frac{1}{2}cx(1-x)^2dx = \int_0^1 c\left(\frac{1}{2}x - x^2 + \frac{1}{2}x^3\right)dx = c\left(\frac{1}{4} - \frac{1}{3} + \frac{1}{8}\right) = \frac{c}{24}. \tag{15}$$

Ifølge (12) skal dette volumet være 1, slik at $c$ må være lik 24 hvis dette skal være en fordeling.

Figur nedenfor gir et bilde av denne fordelingen.

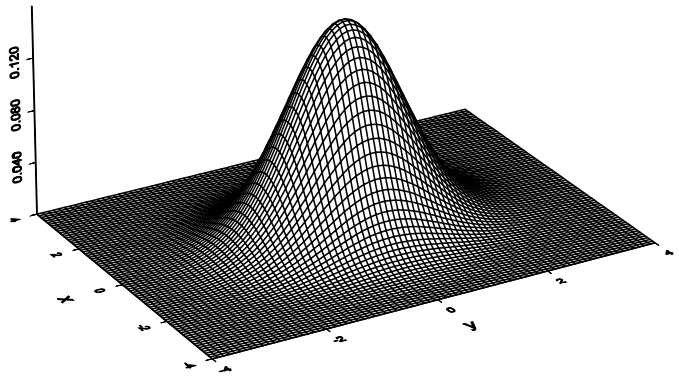



*Eksempel c - to-dimensjonal standard normalfordeling*

I figuren nedenfor ser vi den to-dimensjonale standard normalfordelingen definert ved

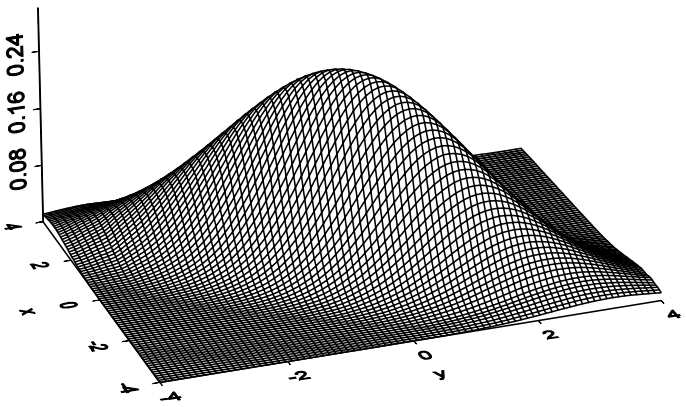
$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \tag{16}$$

definert for alle $(x, y)$.




Vi skal senere behandle den mer generelle binormalfordelingen i et egen notat. Her viser vi en slik standard (forventninger lik null og varianser lik 1) binormalfordeling med korrelasjonskoeffisient $\rho = 0.8$ mellom $X$ og $Y$. Uttrykket for denne er

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/[2(1-\rho^2)]}. \tag{17}$$




# Marginalfordelinger

Ifølge definisjonen (10) kan den kumulative fordelingen til $Y$ uttrykkes som

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(u, v) du dv \tag{18}$$

hvor vi har valgt $u$ og $v$ som integrasjonsvariable for ikke å blande disse sammen med integrasjonsgrensen. Ved å derivere m.h.p. $y$ på begge sider blir venstresiden lik sannsynlighetstettheten til $Y$, mens høyresiden blir integranden i det ytterste integralet innsatt $v = y$ siden vi deriverer m.h.p. øverste integrasjonsgrense. Dette gir

$$f_Y(y) = \int_{-\infty}^{\infty} f(u, y) dx = \int_{-\infty}^{\infty} f(x, y) dx. \tag{19}$$

I dette resonnementet kan vi bytte om $x$ og $y$ og finner da marginalfordelingen til $x$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy. \tag{20}$$

Legge merke til analogien mellom likningene (19) og (20) for kontinuerlige fordelinger og de tilsvarende likningene (4) og (5) for diskrete fordelinger.

*Eksempel a, to-dimensjonal rektangulærfordeling*

Vi finner her marginalfordelingen til X ved å integrere fra 0 til 1 m.h.p. $y$ (fordi sannsynlighetstettheten er null utenfor dette området og at vi derfor ikke får noe bidrag til integralet derfra)

$$f_X(x) = \int_0^1 1 \cdot dy = 1 \tag{21}$$

for alle $x$ i intervallet $[0,1]$, dvs. $X$ er rektangulært fordelt på $[0,1]$.

*Eksempel b*

For å finne marginalfordelingen $f_X(x)$ integrerer vi nå $y$ fra 0 til $(1-x)$ fordi tettheten er null utenfor dette området. Dette gir

$$f_X(x) = \int_0^{1-x} 24xydy = 24x(y^2/2)|_0^{1-x} = 12x(1-x)^2 \tag{22}$$

for $0 \leq x \leq 1$ og null ellers. La oss sjekke at dette faktisk er en fordeling

$$\int_0^1 f_X(x)dx = 12\int_0^1 (x - 2x^2 + x^3)dx = 12\left(\frac{1}{2} - 2\frac{1}{3} + \frac{1}{4}\right) = 1. \tag{23}$$

*Eksempel c - standard normalfordeling*

Marginalfordelingen til $X$ blir her

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{24}$$

som er en-dimensjonal standard normalfordeling.

## Forventninger, varianser og kovarianser

Vi ser nå på den stokastiske variable $Z = g(X, Y)$, definert som en vilkårlig funksjon av $X$ og $Y$. Forventningsverdien til $Z$ er da gitt ved integralet

$$\mathrm{E}Z = \mathrm{E}g(X, Y) = \iint g(x, y) f(x, y) dxdy. \tag{25}$$

Spesielt gjelder dette for $g(X, Y) = Y$ som gir

$$\mathrm{E}Y = \iint y f(x, y) dxdy = \int \left[ y \int f(x, y) dx \right] dy = \int y f_Y(y) dy \tag{26}$$

slik at forventningen til $Y$ blir forventningen i marginalfordelingen til $Y$ (og tilsvarende for $\mathrm{E}X$).

Tilsvarende finnes vi for eksempel variansen til $X$ ved å sette $g(X, Y) = (X - \mathrm{E}X)^2$ og kovariansen mellom $X$ og $Y$ ved å bruke $g(X, Y) = (X - \mathrm{E}X)(Y - \mathrm{E}Y)$. Som for diskrete fordelinger kan vi også her beregne kovariansen ved først å beregne $\mathrm{E}(XY)$ og så trekke fra $\mathrm{E}X\mathrm{E}Y$.

*Eksempel b*

Vi finner her forventningen til $X$ ved å integrere over marginalfordelingen til $X$ som vi allerede har funnet

$$\mathrm{E}X = \int x f_X(x) dx = \int_0^1 12x^2(1 - x)^2 dx = 12 \int_o^1 (x^2 - 2X^2 + x^4) dx = 2/5. \tag{27}$$

Siden fordelingen er symmetrisk i $X$ og $Y$ (vi får samme fordeling om $X$ og $Y$ bytter plass i formelen for fordelingen) blir også $\mathrm{E}Y = 2/5$. Videre finner vi

$$\mathrm{E}(XY) = \iint xyf(x,y)dxdy = 24\int_0^1 [x^2 \int_0^{1-x} y^2 dy]dx = 8\int_0^1 x^2(1-x)^3 dx = 1/10 \tag{28}$$

som gir $\mathrm{cov}(X,Y) = 1/10 - (2/5)^2 = -0.06$.

# Betingede fordelinger

Analogt med likning (7) for diskrete fordelinger defineres den betingede fordelingen til $X$ gitt hendelsen $Y = y$ som

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \tag{29}$$

og tilsvarende for betinging m.h.p. $X$.

*Eksempel b*

Ved å bruke marginalfordelingen vi fant i (22) har vi da at den betingede fordelingen til $Y$ gitt $X$ er

$$f_{Y|X}(y|x) = \frac{24xy}{12x(1-x)^2} = \frac{2y}{(1-x)^2}. \tag{30}$$

Legg merke til at verdiområdet for $Y$ for en gitt $x$ er $[0, 1-x]$. Den betingede fordelingen gitt ved (3) er en vanlig endimensjonal fordeling der den gitte verdien $x$ dukker opp som en parameter. Denne fordelingen har en forventning og varians, som vi da kaller den betingede forventningen og variansen til $Y$ gitt $X = x$. Her blir den betingede forventningen

$$\mathrm{E}(Y|X = x) = \int_0^{1-x} y\frac{2y}{(1-x)^2}dy = \frac{2}{3}(1-x). \tag{31}$$

# Dobbeltforventning

I eksempel b ser vi at forventningen til $Y$ gitt $X = x$ blir en funksjon av $x$, her lik $(2/3)(1 - x)$. Hvis vi nå erstatter den observerte verdien $x$ med den stokastiske variable $X$ fremkommer en ny stokastisk variabel $(2/3)(1 - X)$ som vi kan skrive som $\mathrm{E}(Y|X)$. Vi ser at forventningen til denne funskjonen av $X$ blir

$$\mathrm{E}[\mathrm{E}(Y|X)] = (2/3(1 - \mathrm{E}X) = (2/3)(1 - 2/5) = 2/5 = \mathrm{E}Y. \tag{32}$$

Dette er setningen om dobbeltforventning som generelt sier at $\mathrm{EE}(Y|X) = \mathrm{E}Y$ (og vi kan her bytte om $X$ og $Y$).

Beviset for denne setningen er ganske enkelt

$$\mathrm{E}(Y|X = x) = \int y f_{Y|X}(y|x)dy = \int y \frac{f(x,y)}{f_X(x)}dy. \tag{33}$$

Forventningen til denne funksjonen innsatt $X$ finnes ved å integrere over marginalfordelingen til $X$

$$\mathrm{E}[\mathrm{E}(Y|X)] = \int \left[\int y \frac{f(x,y)}{f_X(x)}dy\right] f_X(x)dx = \iint y f(x,y)dxdy = \mathrm{E}Y. \tag{34}$$

Denne setningen om dobbeltforventning gjelder også for diskrete variable.

# Betinging m.h.p. en hendelse med positiv sannsynlighet

Istedet for å betinge med hensyn på den observerte veridien til en kontinuerlig fordelt variabel, er det ofte aktuelt å betinge med hensyn på en hendelse $A$ med $P(A) > 0$. Vi ser da på en kontinuerlig fordelt stokastisk variabel $X$ som har fordeling som avhenger av om hendelsen $A$ har inntruffet eller ikke. Tilsvarende vil da sannsynligheten for hendelsen $A$ avhenge av verdien til $X$ hvis $X$ er observert. Vi skriver dette som $P(A|X = x)$. La $f(x)$ og

$F(x)$ betegne ubetinget sannsynlighetstetthet og kumulativ fordeling til $X$. Kumulativ fordeling til $X$ gitt hendelsen $A$ defieres da som en vanlig betinget sannsynlighet

$$F_{X|A}(x|A) = P(X \leq x|A) = \frac{P(\{X \leq x\} \cap A)}{P(A)}. \tag{35}$$

Den betingede sannsynlighetstettheten framkommer ved å derivere den kumulative fordelingen

$$f_{X|A}(x|A) = F'_{X|A}(x|A). \tag{36}$$

Disse definisjonene gir opphav til en ny variant av loven om total sannsynlighet

$$P(A) = \int P(A|X = x)f(x)dx \tag{37}$$

og Bayes formel

$$f_{X|A}(x|A) = \frac{P(A|X = x)f(x)}{P(A)} = \frac{P(A|X = x)f(x)}{\int P(A|X = x)f(x)dx}. \tag{38}$$

Vi kan her la $A$ og $X$ bytte roller i formlene. Da blir loven om total sannsynlighet

$$f(x) = f_{X|A}(x|A)P(A) + f_{X|\bar{A}}(x|\bar{A})P(\bar{A}) \tag{39}$$

og Bayes formel

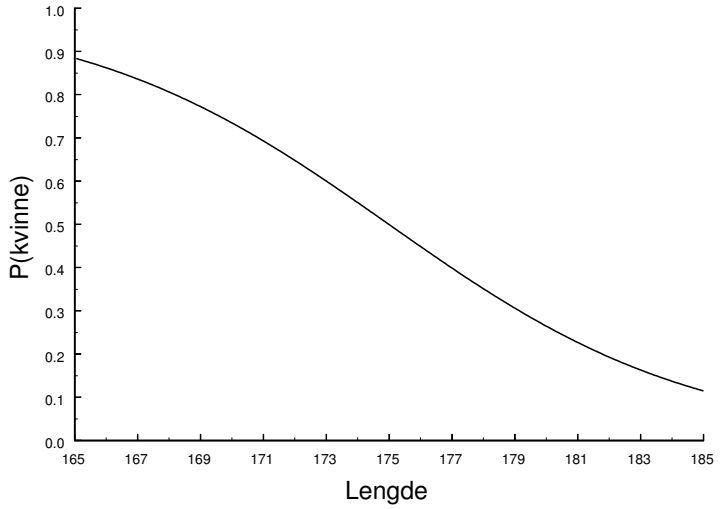
$$P(A|X = x) = \frac{f_{X|A}(x|A)P(A)}{f(x)} \tag{40}$$

hvor nevneren kan uttrykikes ved de betingede fordelingene ved å sette inn fra likning (39).

*Eksempel d - Fordeling av høyden til menn og kvinner*

Anta at høyden til menn er $N(180, 7^2)$ og at høyden til kvinner er $N(170, 7^2)$. Du får vite at en person har høyden $X$. Hva er da sannsynligheten for at det er en kvinne? Vi kan her la $A$ betegne hendelsen {kvinne} og sette inn i likning (40) hvor nevneren regnes ut ved å bruke (39). Ved å multiplisere i teller og nevner med $2 \cdot 7\sqrt{2\pi}$ finner vi

$$P(\text{kvinne}|X = x) = P(A|X = x) = \frac{e^{-(x-170)^2/98}}{e^{-(x-170)^2/98} + e^{-(x-180)^2/98}} \tag{41}$$




*Eksempel e - Fisken i garnet*

Anta at vekten $X$ til individer av en gitt fiskeart har fordeling

$$f(x) = \lambda e^{-\lambda x} \tag{42}$$

for $x \geq 0$. Anta videre at at en fisk som forsøker å passere garnet blir fanget er

$$P(A|X = x) = 1 - e^{-\beta x}. \tag{43}$$

Vi kan nå finne vektfordelingen til de individene som blir fanget ved å sette inn i (38)

$$f_{X|A}(x|A) = \frac{(1 - e^{-\beta x})\lambda e^{-\lambda x}}{\int_0^\infty (1 - e^{-\beta x})\lambda e^{-\lambda x} dx} = \frac{\lambda(\lambda + \beta)}{\beta}(1 - e^{-\beta x})e^{-\lambda x} \tag{44}$$

Forventet vekt til fisk som blir fanget er da

$$\mathrm{E}(X|A) = \int_0^\infty x f_{X|A}(x|A)dx = \frac{2\lambda + \beta}{\lambda(\lambda + \beta)} = \frac{2\lambda + \beta}{\lambda + \beta}\ \mathrm{E}X. \tag{45}$$

*Eksempel f - Poissonblanding*

La $X$ betegne tettheten (forventet antall pr. flateenhet) av individer av en gitt art og anta at denne tettheten varierer stokastisk mellom lokaliteter med kontinuerlig fordeling $F_X(x)$. Anta videre at antall individer $Y$ innenfor en arealenhet ($\mathrm{m}^2$, $\mathrm{km}^2$, ...) på lokalitet med tetthet $X = x$ er Poissonfordelt med parameter $x$, dvs.

$$P(Y = y|X = x) = \frac{x^y}{y!}e^{-x} \tag{46}$$

for $y = 0, 1, \cdots$. Ved å sette $A = \{Y = y\}$ kan vi nå finne den ubetingede fordelingen til $X$, som da kan tolkes som fordelingen til antall individer innen en vilkårlig valgt flateenhet. Likning (37) gir da

$$P(Y = y) = p_Y(y) = \int_0^\infty \frac{x^y}{y!}e^{-x} f_X(x)dx. \tag{47}$$

En diskret fordeling $p_Y(y)$ konstruert på denne måten kalles en Poissonblanding.

Vi kan lett finne forventningen og variansen i denne fordelingen ved å bruke setningen om dobbeltforvening som gjelder også om den ene variable er

diskret og den andre kontinuerlig. Siden $\mathrm{E}(Y|X) = X$ ser vi da at $\mathrm{E}Y = \mathrm{E}\mathrm{E}(Y|X) = \mathrm{E}X$. Vi kan videre bruke samme setning på den stokastiske variable $Y^2$,

$$\mathrm{E}Y^2 = \mathrm{E}\mathrm{E}(Y^2|X) = \mathrm{E}(X + X^2) = \mathrm{E}X + \mathrm{var}(X) + (\mathrm{E}X)^2 \tag{48}$$

som gir

$$\mathrm{var}(Y) = \mathrm{E}Y^2 - (\mathrm{E}Y)^2 = \mathrm{E}Y + \mathrm{var}(X) \tag{49}$$

hvor vi har benyttet at $\mathrm{E}Y = \mathrm{E}X$. Ved å blande Poissonfordelinger på denne måten ser vi at fordelingen til $X$ får større varians enn forventning, mens disse to parametrene er like for Poissonfordelte variable. Vi sier da at vi har en fordeling med overdispersjon relativt til Poissonfordelingen.

*Eksempel g - 'Screening' for å oppdage brystkreft*

La $X$ betegne den tiden det tar fra en en kreftsvulst i brystet begynner å vokse til kvinnen selv oppdager den og kontakter lege. Det viser seg at denne tiden varierer mye fra kvinne til kvinne siden noen svulster vokser fort og andre langsomt. La $f_X(x)$ betegne sannsynlighetstettheten til $X$. Vi antar at en gruppe kvinner i aktuell alder blir innkalt til kontroll. En slik kontroll kan da avsløre svulsten på et tidligere tidspunkt. Anta at sannsynligheten for at en kvinne med tid $x$ blir innkalt til kontroll etter at svulsten har startet å vokse men før hun selv oppdager den er tilnærmet $\varepsilon x$, altså proportionalt med tidsintervallets lengde. Vi skriver dette

$$P(A|X = x) = \varepsilon x \tag{50}$$

og antar da at $x$ ikke kan ta så store verdier at sannsynligheten $\varepsilon x$ overskrider 1. Fordelingen til $X$ gitt $A$ er da gitt ved likning (38)

$$f_{X|A}(x|A) = \varepsilon x f_X(x) / \left( \int_0^\infty \varepsilon x f_X(x) dx \right) = x f_X(x)/\mu \tag{51}$$

hvor $\mu = \mathrm{E}X$, altså midlere tid til svulsten oppdages av kvinnen selv. Forventningen i denne fordelingen blir

$$\mathrm{E}(X|A) = \int_0^\infty x F_{X|A}(x)dx = \int_=^\infty x^2 f_X(x)/\mu dx = \mathrm{E}X^2/\mathrm{E}X. \tag{52}$$

Vi ser nå på de tilfellene som blir oppdaget ved kontrollen og lar $Y$ betegne hvor lenge svulster som oppdages på denne måten har vokst. En rimelig anttagelse er da at $Y$ betinget m.h.p. $X = x$ er rektangulært fordelt mellom 0 og $x$, dvs.

$$F_{Y|X}(y|x) = 1/x \text{ for } 0 \leq y \leq x. \tag{53}$$

Den ubetingede fordelingen til $Y$ gitt hendelsen $A$ blir da

$$f_{Y|A}(y) = \int_y^\infty \frac{1}{x}\frac{x f_X(x)}{\mu}dx = (1 - F_X(y))/\mu. \tag{54}$$

Forventet alder til de svulstene som oppdages i kontrollen, dvs. E$Y$, finnes lettest ved å bruke setningen om dobbeltforventning. Vi bruker da setningen i utfallsrommet $A$, dvs. alle hendelser betinges m.h.p. $A$. Siden $Y$ er rektangulært fordelt på $[0, x]$ gitt at $X = x$, har vi at $\mathrm{E}(Y|X, A) = X/2$. Herav følger at

$$\mathrm{E}Y = \mathrm{E}\mathrm{E}(Y|X, A) = \mathrm{E}(X/2|A) = \frac{1}{2}\mathrm{E}(X|A) = \frac{1}{2}\mathrm{E}X^2/\mathrm{E}X. \tag{55}$$

Vi ser nå at forventet alder til de svulstene som oppdages i kontrollen er større enn den forventede alderen de svulstene som oppdages av kvinnene selv dersom $\frac{1}{2}\mathrm{E}X^2/\mathrm{E}X > \mathrm{E}X$. Denne betingelsen kan skrives som $\frac{1}{2}[\mathrm{E}X^2 - (\mathrm{E}X)^2] > \frac{1}{2}(\mathrm{E}X)^2$, eller

$$\mathrm{var}(X)/(\mathrm{E}X)^2 > 1. \tag{56}$$

Variasjonskoeffisienten til en positiv stokastisk variabel defineres som forholdet mellom standardavviket og forventningen og betegnes ofte med $C$. Vi ser da at forventet alder til svulstene som oppdages i kontrollen er større enn forventet lengde til svulstene som kvinnene selv oppdager hvis $C > 1$.

Veksten til slike svulster varierer veldig mellom individer, så fordelingen til $X$ kan være en veldig skjev fordeling med $C > 1$.

Dette resultatet kan virke paradoksalt fordi at alle $Y$-verdier som observeres nødvendigvis er mindre enn kvinnens $X$-verdi. Forklaringen ligger i at kvinnene med voksende svulst ikke observeres med samme sannsynlighet, og at fordelingen til $X$ derfor i praksis blir fordelingen gitt ved likning (51) og ikke den ubetingede fordelingen $f_X(x)$.

Forventningen til $Y$ kan alternativt finnes ved å integrere over fordelingen (54). For å løse integralet må man bruke delvis integrasjon.