Binomisk fordeling tilnærmet med normalfordeling (6.5)

X = "antall suksesser i n forsøk" er en sum (la "suksess"=1 og "ikke suksess"=0) og kan dermed i følge SGT tilnærmes med normalfordeling. Når $X \sim B(n,p)$ husker vi at $\mathrm{E}(X) = np$ og $\mathrm{Var}(X) = np(1-p)$, og vi får dermed at

$$Z = \frac{X - \overbrace{np}^{E(X)}}{\underbrace{\sqrt{np(1-p)}}_{Var(X)}} \approx N(0,1)$$

OK tilnærming når $np > 5$ og $n(1-p) > 5$. Se figurene 6.22-6.25 i boka. Evt. $np(1-p) > 5$
Var$(X) > 5$

$P(a \leq X \leq b)$

$$\approx P\Big(\frac{a - 0.5 - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\Big)$$

$$= P\Big(Z \leq \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\Big) - P\Big(Z \leq \frac{a - 0.5 - np}{\sqrt{np(1-p)}}\Big)$$

Her er -0.5 i nedre grense og +0.5 i øvre grense korreksjoner vi gjør for å få en bedre tilnærming (se fig. 6.23).

ch is approximately equal to the area of the shaded region under the norma ve between the two ordinates $x_1 = 3.5$ and $x_2 = 4.5$ in Figure 6.23. Converting z values, we have

$$z_1 = \frac{3.5 - 6}{1.897} = -1.32 \quad \text{and} \quad z_2 = \frac{4.5 - 6}{1.897} = -0.79.$$

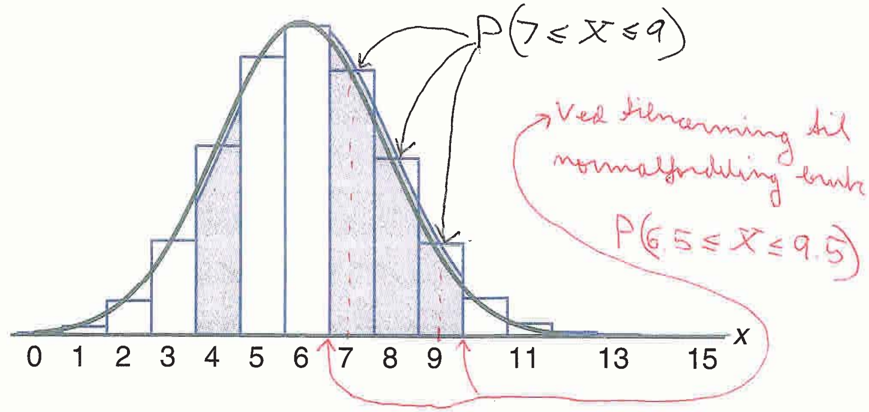


Figure 6.23: Normal approximation of $b(x; 15, 0.4)$ and $\sum_{x=7}^{9} b(x; 15, 0.4)$.

X is a binomial random variable and Z a standard normal variable, then

$$P(X = 4) = b(4; 15, 0.4) \approx P(-1.32 < Z < -0.79)$$
$$= P(Z < -0.79) - P(Z < -1.32) = 0.2148 - 0.0934 = 0.1214.$$

s agrees very closely with the exact value of 0.1268.

The normal approximation is most useful in calculating binomial sums for large ıes of n. Referring to Figure 6.23, we might be interested in the probability t X assumes a value from 7 to 9 inclusive. The exact probability is given by

$$P(7 \leq X \leq 9) = \sum_{x=0}^{9} b(x; 15, 0.4) - \sum_{x=0}^{6} b(x; 15, 0.4)$$
$$= 0.9662 - 0.6098 = 0.3564,$$

ch is equal to the sum of the areas of the rectangles with bases centered a 7, 8, and 9. For the normal approximation we find the area of the shaded on under the curve between the ordinates $x_1 = 6.5$ and $x_2 = 9.5$ in Figure 6.23 corresponding z values are

$$z_1 = \frac{6.5 - 6}{} = 0.26 \quad \text{and} \quad z_2 = \frac{9.5 - 6}{} = 1.85$$

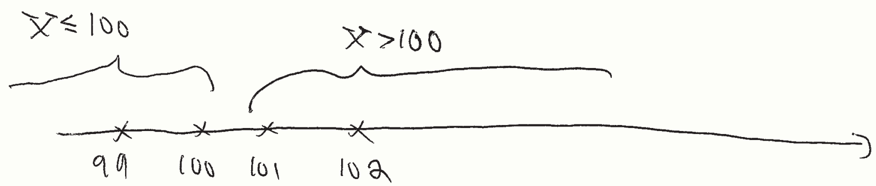
Poissonfordeling tilnærmet med normalfordeling

Også Poisson-fordelte variable kan tilnærmes med normalfordeling. Dersom $\lambda t > 15$ er

$$
\begin{aligned}
P(a \le X \le b) & \\
\approx \; & P\Big(\frac{a-0.5-\overset{E(X)}{\lambda t}}{\underset{Var(X)}{\sqrt{\lambda t}}} \le Z \le \frac{b+0.5-\lambda t}{\sqrt{\lambda t}}\Big)
\end{aligned}
$$

en god tilnærming som kan brukes til å forenkle beregninger. (Husk at $E(X) = Var(X) = \mu = \lambda t$ i Poisson-fordeling.)

Eksempel: Antall visninger av en webside per dag er Poisson-fordelt med forventing 92. → $\lambda = 92$ Hva er sannsynligheten for at siden har mer enn 100 visninger i løpet av en dag? ($t=1$) Hva er sannsynligheten for at siden har mer enn 700 visninger i løpet av en uke? ($t=7$)



La X=”antall visninger per dag” og Y=”antall visninger per uke”. Vi har da at X ~Poisson(92) og Y ~Poisson($\underset{\lambda \cdot t}{92 \cdot 7}$).

$$
\begin{aligned}
P(X > 100) &= 1 - P(X \le 100) \\
&\approx 1 - P\Big(Z \le \frac{100 + 0.5 - \overset{E(X)}{92}}{\underset{Var(X)}{\sqrt{92}}}\Big) \\
&= 1 - P(Z \le 0.89) \\
&= 1 - 0.8133 = \underline{\underline{0.1867}} \\
P(Y > 700) &= 1 - P(Y \le 700) \\
&\approx 1 - P\Big(Z \le \frac{700 + 0.5 - 92 \cdot 7}{\sqrt{92 \cdot 7}}\Big) \\
&= 1 - P(Z \le 2.23) \\
&= 1 - 0.9871 = \underline{\underline{0.0129}} \quad \square
\end{aligned}
$$

Gammafordeling (6.6-6.7)

Gammafordelingen er en fleksibel fordelingsklasse for $x \geq 0$. Gammafordeling brukes bl.a. til å beskrive levetider/funksjonstider både for tekniske systemer og i biologi. Flere mye brukte fordelinger er spesialtilfeller av gammafordelingen.

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0$$

$E(X) = \alpha\beta$ og $\mathrm{Var}(X) = \alpha\beta^2$. $\Gamma(\alpha)$ er gammafunksjonen. Se formelsamlingen side 33. Merk spesielt at $\Gamma(n+1) = n!$ dersom n er et heltall.

$\Gamma(n) = (n-1)!$ $\qquad \Gamma(x) = (x-1)\Gamma(x-1)$

$\Gamma(12) = 11!$ $\qquad \Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{1}{2}\sqrt{\pi}$

Eksempel: Levetiden til en bestemt type insekt i måneder er beskrevet ved en gammafordeling med parametre $\alpha = 2$ og $\beta = 0.5$. Hva er sannsynligheten for at et insekt levere i mer enn 2 måneder?

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - \int_0^2 \frac{1}{0.5^2\Gamma(2)} x^{2-1} e^{-x/0.5} dx \\ &= 1 - 4\int_0^2 x e^{-2x} dx \\ &= 1 - 4 \cdot 0.227 = \underline{\underline{0.092}} \quad \Box \end{aligned}$$

$$= \int_2^\infty f(x)\,dx = \int_2^\infty \frac{1}{0.5^2 \underbrace{\Gamma(2)}_{=1!=1}} x^{2-1} e^{-x/0.5}\,dx$$

$$= \int_2^\infty 4 \cdot x e^{-2x}\,dx = \underline{\underline{0.092}}$$

Eksponensialfordeling (6.6-6.7)

Et viktig spesialtilfelle av gammafordelingen er eksponensialfordelingen som vi får ved å sette $\alpha = 1$ i gammafordelingen:

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x \geq 0,$$

$E(X) = \beta$ og $\mathrm{Var}(X) = \beta^2$ (se øving 2).

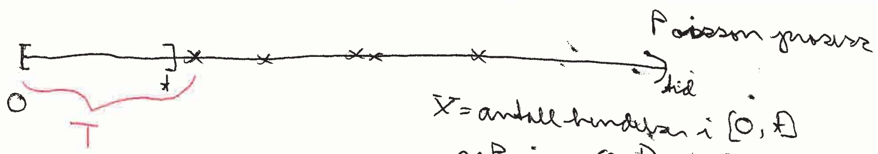
$\hookrightarrow \int_{-\infty}^{\infty} x f(x)\,dx = \ldots = \beta$

Merk: Ofte brukes parameteren $\lambda = 1/\beta$.

$f(x) = \lambda e^{-\lambda x}$, $x \geq 0$, $E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

Eksponensialfordelingen brukes bl.a. til å modellere funksjonstider (bl.a. til elektroniske komponenter), tider til og mellom hendelser, etc.

Videre har eksponensialfordelingen en del viktige relasjoner til Poisson-prosser.



X = antall hendelser i $[0, t)$

$\sim$ Poisson$(\lambda \cdot t)$

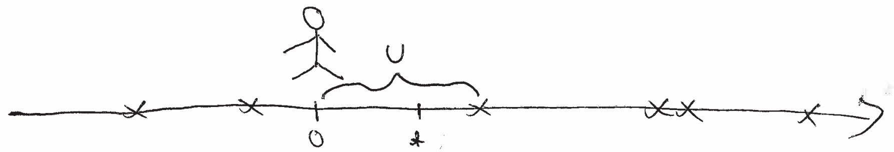
La T være tid til første hendelse i en Poisson-prosess. Vi har da at: $f(x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$

$$\begin{aligned} P(T > t) &= P(\text{ingen hendelser i } [0,t]) \\ &= p(0; \lambda t) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t} \end{aligned}$$

Dvs $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$, og dermed blir tettheten for tid til første feil $f(t) = F'(t) = \lambda e^{-\lambda t}$ som er en eksponsialfordeling (med $\lambda = 1/\beta$). $f(t) = F'(t)$

Dvs tid til første hendelse i en Poisson-prosess er eksponensialfordelt. Det kan på tilsvarende måte også vises at tid *mellom* hendelser i en Poisson-prosess er eksponensialfordelt (f.eks. tid mellom visninger av en webside, tid mellom utrykninger fra en brannstasjon osv.)

Det kan også vises at tid til $k$te hendelse i en Poisson-prosess er gammafordelt med parametre $\alpha = k$ og $\beta = 1/\lambda$. Vi skal se litt mer på disse tingene senere.



Eksempel: Bussproblemet. Anta at ankomsten av busser til en bussholdeplass kan modelleres som en Poisson-prosess med parameter λ. Tiden mellom hver gang det kommer en buss er da eksponensialfordelt med forventningsverdi $1/\lambda$.

Du ankommer bussholdeplassen. Hva er forventet tid du må vente før det kommer en buss? La U være ventetiden fra du ankommer til det kommer en buss.

$$\begin{aligned} P(U > t) &= P(\text{ingen busser i } [0,t]) \\ &= \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t} \quad \Rightarrow U \sim \text{eksp}(\lambda) \end{aligned}$$

$F_U(t) = P(U \le t) = 1 - e^{-\lambda t}$

$f_U(t) = \lambda e^{-\lambda t}$

dvs $\mathrm{E}(U) = 1/\lambda$ - samme som for tid mellom busser!!

Hvorfor? Når vi ankommer til bussholdeplassen har vi større sannsynlighet for å treffe et stort tidsintervall mellom busser enn et lite. □

χ^2-fordeling (6.8)

Viktig spesialtilfelle av gammafordelingen der $\alpha = \nu/2$ og $\beta = 2$ der ν er et positivt heltall. Brukes mye i statistisk teori, bl.a. ifbm tester/konfidensintervall for varians.

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x \ge 0,$$

$$\nu = 1, 2, \ldots$$

$\mathrm{E}(X) = \nu$ og $\mathrm{Var}(X) = 2\nu$.

$\nu = 2$

$$f(x) = \frac{1}{2^{2/2}\Gamma(2/2)} x^{2/2-1} e^{-x/2}$$

$$= \frac{1}{2} e^{-x/2}$$

$\ln(X) \sim N(\mu, \sigma^2)$

$X \sim \text{lognormal}(\mu, \sigma^2)$

Lognormal-fordeling (6.9)

X er lognormal-fordelt dersom $\ln(X)$ er normalfordelt. Dersom i tillegg $\mathrm{E}(\ln(X)) = \mu$ og $\mathrm{Var}(\ln(X)) = \sigma^2$ er

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{1}{2} \frac{(\ln(x) - \mu)^2}{\sigma^2} \right), \quad x > 0,$$

$\mathrm{E}(X) = e^{\mu + \sigma^2/2}$ og $\mathrm{Var}(X) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$.
Lognormalfordeling brukes bl.a. som modell for konsentrasjoner, levetider, utmatting, etc

Eksempel: Konsentrasjonen (målt i p.p.m.) av et forurensende stoff i utslippsvannet fra en kjemisk fabrikk er lognormalfordelt med parametre $\mu = 3.2$ og $\sigma = 1$. Hva er sannsynligheten for at konsentrasjonen overstiger 8?

$$\begin{aligned}
P(X > 8) &= 1 - P(X \le 8) \\
&= 1 - P(\ln(X) \le \ln(8)) \\
&= 1 - P(\frac{\ln(X) - 3.2}{1} \le \frac{\ln(8) - 3.2}{1}) \\
&= 1 - P(Z \le -1.12) \\
&= 1 - 0.1314 = \underline{\underline{0.8686}} \quad \square
\end{aligned}$$

$F(x) = P(X \le x) = \int_0^x f(v)\,dv = \int_0^x \alpha\beta v^{\beta-1} e^{-\alpha v^\beta} dv = \left[-e^{-\alpha v^\beta}\right]_0^x = -e^{-\alpha x^\beta} - (-1)$

$= \underline{\underline{1 - e^{-\alpha x^\beta}}}$

Weibull-fordeling (6.10)

Weibull-fordeling blir bl.a. brukt til å modellere levetider/funksjonstider til komponenter og system. Enklere å regne med en gamma-fordelingen.

$$f(x) = \underbrace{\alpha\beta x^{\beta-1}}_{u'} \underbrace{e^{-\alpha x^\beta}}_{u}, \quad x \ge 0,$$

$\mathrm{E}(X) = \alpha^{-1/\beta}\Gamma(1 + \frac{1}{\beta})$ og

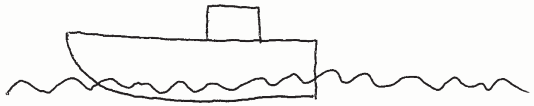
$\mathrm{Var}(X) = \alpha^{-2/\beta}\left[\Gamma(1 + \frac{2}{\beta}) - \Gamma(1 + \frac{1}{\beta})^2\right]$.

Merk: $\beta = 1$ gir eksponensialfordeling.

Eksempel: Oppetiden (i dager) for en server kan modelleres ved en Weibull-fordeling med $\alpha = 0.25$ og $\beta = 0.5$. Hva er sannsynligheten for at serveren har en oppetid på mindre enn 30 dager?

$$\begin{aligned}
P(X < 30) &= \int_0^{30} 0.25 \cdot 0.5 x^{0.5-1} e^{-0.25x^{0.5}} dx \\
&= [-e^{-0.25x^{0.5}}]_0^{30} \\
&= -e^{-0.25 \cdot 30^{0.5}} - (-1) = \underline{\underline{0.746}} \quad \square
\end{aligned}$$

Eksempel Bølgehøyde



La X være høyden til en tilfeldig bølge.

Vanlige modell for fordelingen til bølgehøyde under noenlunde stabile bølgeforhold er fordeling gitt ved sannsz. tetthet

$$f_X(x) = \frac{2x}{\theta} e^{-x^2/\theta} \quad , \quad x > 0$$

Kalles Rayleigh - fordeling.

Merk! Spesialtilfelle av Weibull med $\beta = 2$ og $\alpha = \frac{1}{\theta}$

Vet da at:

$$E(X) = \alpha^{-1/\beta}\Gamma\left(1+\frac{1}{\beta}\right) = \theta^{1/2}\underbrace{\Gamma\left(1+\frac{1}{2}\right)}_{\frac{1}{2}\sqrt{\pi}} = \frac{1}{2}\sqrt{\theta\pi}$$

Anta at vi har bølgeforhold som gir $\theta = 2.5$.

Da vil: $E(X) = \frac{1}{2}\sqrt{2.5 \cdot \pi} = 1.40$

$$P(X>2) = \int_2^\infty f_X(x)\,dt = \int_2^\infty \frac{2x}{2.5} e^{-x^2/2.5}\,dt$$

$$= \left[-e^{-x^2/2.5}\right]_2^\infty = 0 - \left(-e^{-2^2/2.5}\right) = 0.20$$