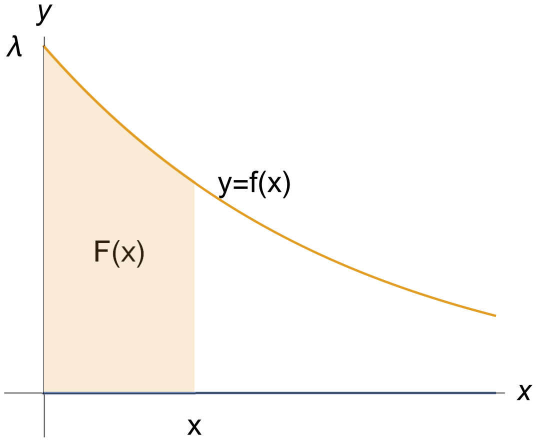
1 Kontinuerliga fördelningar

Till skillnad från en diskret stokastisk variabel, som bara kan anta ett antal värden (ändligt eller uppräkneligt oändligt) så kan en kontinuerlig stokastisk variabel anta alla värden i ett intervall. Exempel på storheter som beskrivs med kontinuerlig stokastisk variabel är



tid, längd, vikt, area, energi, kraft, kraftmoment

En kontinuerlig stokastisk variabel har i allmänhet **väntevärde** och **varians**. Vi definierar dessa begrepp senare men kommer ändå att presentera dessa samtidigt med presentationen av respektive fördelning. En kontinuerlig stokastisk variabel har en frekvensfunktion med egenskaper som i (3). Sannolikheten för en händelse ges av arean under frekvensfunktionen. Mer exakt definierar vi följande (Se figur 1).

Definition 1
Om det finns en funktion f sådan att $f(x) \geq 0$ för alla x och

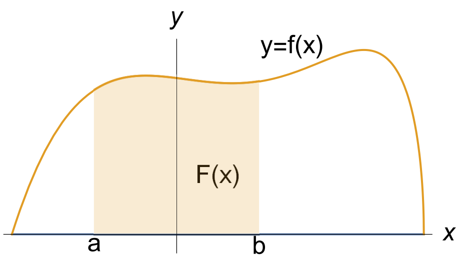
$$\int_{-\infty}^{\infty} f(t)dt = 1 \tag{1}$$

så är f en sannolikhets- eller frekvensfunktion. Fördelningsfunktionen definieras som

$$F(x) := P(\xi \leq x) = \int_{-\infty}^{x} f(t)dt \tag{2}$$

Sats 1
Låt ξ vara en kontinuerlig stokastisk variabel med frekvensfunktion f och fördelningsfunktion F. Då gäller följande

$$
\begin{aligned}
&a) \quad F'(x) = f(x) \text{ (utom möjligen i vissa skarvpunkter)} \\
&b) \quad P(a < \xi \leq b) = \int_a^b f(x)dx = F(b) - F(a) \\
&c) \quad P(\xi > x) = \int_x^\infty f(t)dt = 1 - F(x) \\
&d) \quad P(\xi = x) = 0 \text{ för alla } x
\end{aligned} \tag{3}
$$



Figur 1: *Illustration av frekvens- och fördelningsfunktion.*

I figur 1 illustreras sannolikheten $P(\xi \leq 2.2)$, där kurvan ges av $y = f(x)$ och f är ξ:s frekvensfunktion. Enligt integralkalkylens huvudsats är arean $P(a < \xi \leq b) = F(b) - F(b) = \int_a^b f(x)dx$ där F är fördelningsfunktionen till ξ.

Ex 4. 1 Feltoleransen ζ (i mm) för en bults diameter är givet av frekvensfunktionen

$$f(t) = \begin{cases} A(1 - 4t^2), & -0.5 \leq t \leq 0.5 \\ 0, & t < -0.5 \text{ eller } t > 0.5 \end{cases}$$

Vad är sannolikheten att feltoleransen är till sitt belopp mindre än 0.2?

Lösning

Bestäm konstanten A så att vi får $f(t)$ blir en frekvensfunktion. Vi skall alltså beräkna konstanten A så att

$$A \int_{-\infty}^{\infty} f(t)dt = 1 \text{ d.v.s. } A \int_{-0.5}^{0.5} (1 - 4t^2)dt = 1$$

Då denna funktion är jämn får vi att

$$2A \int_0^{1/2} (1 - 4t^2)dt = 2A \left[t - \frac{4t^3}{3}\right]_0^{1/2} = 2A \left(\frac{1}{2} - \frac{4}{24}\right) =$$

$$= 1 \Leftrightarrow A = \frac{3}{2}$$

Därefter integrerar vi

$$\frac{3}{2}\int_{-0.2}^{0.2}(1-4t^2)dt = 3\left[t-\frac{4t^3}{3}\right]_0^{0.2} \approx 0.568$$

Sannolikheten är alltså 56.8% att feltoleransen ligger mellan -0.2 och 0.2.

■

1.1 Rektangelfördelning

Ex 4. 2 När man väntar på bussen och endast vet att bussen kommer med 10 minuters mellanrum kan väntetiden anta alla värden mellan 0 och 10 minuter. Väntetiden är alltså en kontinuerlig stokastisk variabel, säg ξ. Sannolikheten för att bussen kommer mellan 3:e och 4.e minuten som man väntar bör rimligen vara $\frac{4-3}{10} = 0.1$ men också

$$P(\xi \leq 4) - P(\xi \leq 3) = \int_3^4 f(x)dx = [F(x)]_3^4$$

där f är frekvensfunktionen och F är fördelningsfunktionen. Det är då rimligt att

$$P(\xi \leq x) = F(x) = \frac{x}{10}, \quad \text{om } 0 \leq x \leq 10$$

Om $x < 0$ och $x > 10$ är det rimligt att sätta sannolikheten $= 0$ respektive 1. Sammanfattningsvis så är alltså fördelningsfunktionen Frekvensfunktionen erhålls genom att derivera denna funktion.

■

Om ξ är en kontinuerlig stokastisk variabel vars frekvensfunktion är konstant i ett intervall $[a, b]$ och 0 utanför detta intervall säger man att ξ är *rektangelfördelad på intervallet* $[a, b]$. Vi skriver $\xi \in Rab$. Konstanterna a och b är parametrar i fördelningen. Vi ser att

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x > b \end{cases} \tag{4}$$

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases} \tag{5}$$

Ex 4. 3 Betrakta rektangelfördelningen i

(a) Vad är sannolikheten att väntetiden är längre än 7 minuter?

(b) Vad är förväntad väntetid?

(c) En person anländer till hållplatsen och får veta att ena annan person redan har väntat 3 minuter. Vad är sannnolikheteten, att den nyanlända personen bar behöver vänta högst två minuter?

Lösning

(a) Denna sannolikhet kan vi uttrycka $P(\xi > 7)$ med samma ξ som i Vi får att
$$P(\xi > 7) = \int_7^\infty f(x)dx = \int_7^{10} \frac{dx}{10} = \frac{10-7}{10}$$
Alternativt kan man lösa detta med fördelningsfunktionen.
$$P(\xi > 7) = \lim_{x\to\infty} F(x) - F(7) = 1 - F(7) = 1 - \frac{7}{10} = 0.3$$

(b) Förväntad väntetid är givetvis 5 minuter. Vi skall längre fram definiera förväntat värde eller väntevärde.

(c) Sanolikheten för den nyanlända att vänta mindre än 2 minuter, givet att den första har väntat mer än 3 minuter är
$$P(\xi \leq 5|\xi \geq 3) = \frac{P(\xi \geq 3 \wedge \xi \leq 5)}{P(\xi \geq 3)} = \frac{0.2}{0.7} = 0.28 (= 28\%).$$

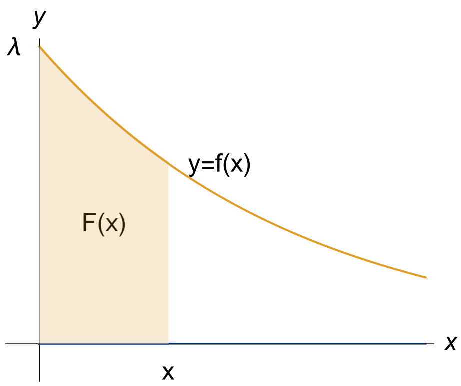
■

1.2 Exponentialfördelning

Om ξ är en kontinuerlig stokastisk variabel med frekvensfunktionen
$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$
där $\lambda > 0$ är en konstant, så säger vi att ξ *är exponentialfördelad med parameter* λ. Vi skriver $\xi \in \mathrm{Exp}(\lambda)$. Frekvensfunktionens utseende för några olika värden på λ framgår ur figur 2. Fördelningsfunktionen fås genom att integrera frekvensfunktionen. För $t < 0$ är $F(t) = 0$. För $t > 0$ får vi
$$F(t) = \int_{-\infty}^t f(x)dx = \int_0^t \lambda e^{-\lambda x}dx = \left[-e^{-\lambda x}\right]_0^t = -(e^{-\lambda t} - 1) = 1 - e^{-\lambda t}$$

Fördelningsfunktionen för ξ är alltså
$$F(t) = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-\lambda t}, & t > 0 \end{cases} \tag{6}$$
Vi ser även att om $t \to \infty$ så går $F(t)$ mot 1. Dessutom gäller att $F(t) \to 0$, då $t \to -\infty$ (eftersom $F(t) = 0$ för alla $t < 0$.) Dessa egenskaper gäller allmänt för alla fördelningsfunktioner.

Ex 4. 4 Ett relä har en livslängd som är exponentialfördelad med $\lambda = 0.004$ (h^{-1}).



Figur 2: Frekvensfunktion för exponentialfördelad variabel.

(a) Beräkna sannolikheten att livslängden överskrider 150 h.

(b) Beräkna sannolikheten att att livslängden överskrider 200 h, om livslängden har överskridit 50 h.

Lösning

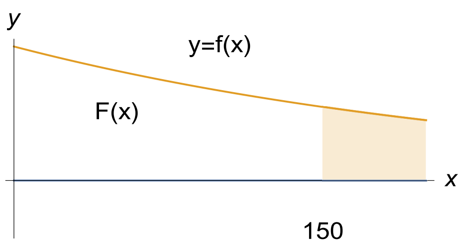
Vi låter ξ vara reläets livslängd. a)

$$P(\xi \geq 150) = 1 - F(150) = e^{-150 \cdot 0.004} \approx 0.55$$

b)

$$P(\xi \geq 200 | \xi \geq 50) = \frac{e^{-200 \cdot 0.004}}{e^{-50 \cdot 0.004}} =$$

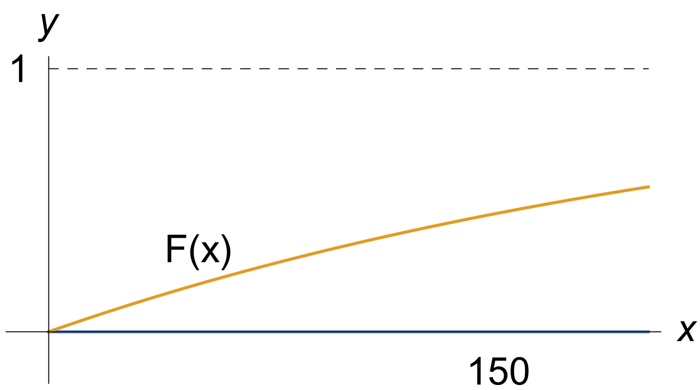
$$= e^{(50-200) \cdot 0.004} = e^{-150 \cdot 0.004} \approx 0.55$$



Figur 3: Sannolikheten $P(\xi \geq 150) = e^{-0.004 \cdot 500} \approx 0.55$ är arean av ytan markerad med orange (4).

■

Kommentarer



Figur 4: Fördelningsfunktionen $F(x) = 1 - e^{-0.004x}$ för $x \geq 0$ från **??**explivs. Sannolikheten för $P(\xi \leq 150)$ är markerad på $y-$axeln.

i) Lösningen i b) visar att *om* livlängden överskridit 50 h så är sannolikheten att den överskrider ytterligare 150 h densamma som i a), d.v.s. relät har inte åldrats under de första 50 timmarna! Exponentialfördelningen beskriver livslängden på enheter/produkter som inte åldras i fysisk mening, såsom vissa elektroniska komponenter.

ii) För $x \geq 0$ så ser grafen till (6) med $\lambda = 0.004$ ut som i figur 4.

1.3 Weibullfördelning

Denna fördelning används för att beskriva hållfasthet (utmattning)och livslängd.

En stokastisk variabel ξ är *Weibull-fördelad med parametrar* $\alpha > 0$ *och* $\beta \geq 1$ om den har fördelningsfunktionen

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x^{\beta}/\alpha}, & x \geq 0 \end{cases} \tag{7}$$

Den är en generalisering av exponentialfördelningen med $\beta = 1$ och $\lambda = 1/a$. med $\alpha = 3$ och $\beta = 2$, $y = \dfrac{2\,x}{3} e^{-\frac{x^2}{3}}$.

1.4 $\Gamma-$fördelningen

En fördelning där frekvensfunktionen är

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{\lambda^n}{(n-1)!} \cdot x^{n-1} \cdot e^{-\lambda x} x \geq 0 \end{cases} \tag{8}$$

kallas $\Gamma-$ (“gamma-”) fördelad av ordning n. Att ξ är gammafördelad skrivs $\xi \in \Gamma(n, \lambda)$. Fördelningen beror på två parametrar n och λ. Om $n = 1$ får vi exponentialfördelningen.

Ex 4. **5** Inom kvantfysik förekommer frekvensfunktioner såsom $f(x) = A \sin^2 kx, \quad 0 \leq x \leq b$ d.v.s. ett maximum ($f(x) = 0$ f.ö.) . Man tror nu kanske att en frekvensfunktion inte kan ha fler än en “topp”, . Men b kan vara sådant att kurvan får två eller fler maxima.

■

Ex 4. 6 Avgör vilka av följande funktioner är frekvensfunktioner.

(a)

$$f(x) = \begin{cases} xe^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(b)

$$f(x) = \begin{cases} 3x^2 + 2x - 1 & 0 \leq x \leq 1 \\ 0 & x < 0 \text{ eller } x > 1 \end{cases}$$

$$f(x) = \begin{cases} \dfrac{\sin x}{2} & 0 \leq x \leq \pi \\ 0 & x < 0 \text{ eller } x > \pi \end{cases}$$

■

Ex 4. 7 Avgör vilka funktioner som är fördelningsfunktioner.

$$F(x) = \begin{cases} 2\,x^2 - x^3 & 0 \leq x \leq 1 \\ 0 & x < 0 \\ 1 & x > 1 \end{cases}$$

■

1.5 Weibullfördelning och livslängd

Ex 4. 8 Livslängden på ett givet rullager till vindkraftverk gäller att dess livslängd t i år Hart fördelningsfunktionen, (en Weibullfördelad fördelningsfunktion)

$$F(t) = \begin{cases} 0, & \text{om } t < 0 \\ \\ 1 - e^{-\sqrt{t}}, & \text{om } t \geq 0. \end{cases}$$

(a) Beräkna sannolikheten att livslängden är länge än 1 år.

(b) Beräkna sannolikheten att livslängden ≥ 2 år om livslängden ≥ 1 år.

(c) Beräkna sannolikheten att livslängden ≥ 3 år om livslängden ≥ 2 år.

(d) Kommentera (b) och (c).

Lösning

(a) Sannolikheten att livslängden är längre än 10 år är $1 - F(10.0) = 0.37$.

(b) Denna sannolikhet är en betingad och är

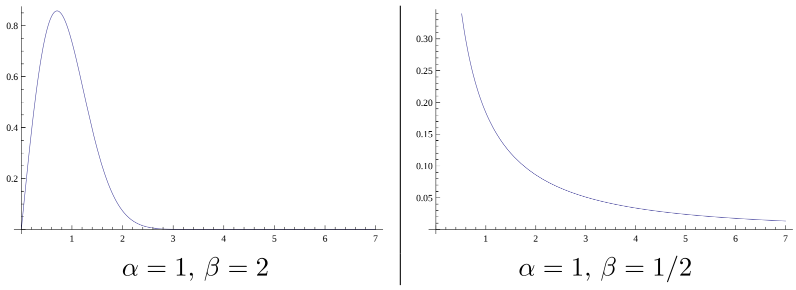
$$\frac{P(\xi > 2)}{P(\xi > 1)} = 0.83.$$

(c) Även denna sannolikhet är betingad och är

$$\frac{P(\xi > 3)}{P(\xi > 2)} = 0.88.$$

(d) Vi ser att sannolikheten för att livslängden ökar i följande mening: Sannolikheten att livslängden är att "leva" minst ett år till är högre efter 2 än efter 1 år. Man kan uttrycka det som att rullagret är "yngre" vid två års ålder än vid ett års ålder. Detta är typiskt material/produkter, som har "barnsjukdomar".

■



$\alpha = 1$, $\beta = 2$ $\alpha = 1$, $\beta = 1/2$

1.6 Normalfördelning

Normalfördelningen är den mest kända av alla kontinuerliga fördelningar. Vi kommer att behandla denna i senare varför vi endast tar upp den kortfattat här. Om den stokastiska variabeln ξ har frekvensfunktionen

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty, \tag{9}$$

säger vi att ξ *är normalfördelad med parametrar μ och σ*. Parametrarna μ och σ är givna konstanter sådana att $-\infty < \mu < \infty$ och $\sigma > 0$. Vi skriver $\xi \in \mathrm{N}(\mu, \sigma)$. Vi ser att frekvensfunktionen är symmetrisk kring μ, som är fördelningens väntevärde[1]. Parametern μ kan sägas ange läget av fördelningen. Parametern σ ökar då spridningen ökar. σ är fördelningens standardavvikelse. Bl.a. används den vid mätning (mätfel). När man betraktar större summor av stokastiska variabler kan dessa approximeras med en normalfördelning. Statistisk fysik (termodynamik) är ett ämnesområde där approximation med normafördelning används. Den kurva som är mest koncentrerad kring väntevärdet 10 i figur 5

Figur 5: Två av normalfördelningens frekvensfunktioner med samma väntevärde $\mu = 10$ och med $\sigma_1 = 1$ och $\sigma_2 = 2$.

har minst standardavvikelse. För att beräkna $P(\xi \leq a)$ om $\xi \in \mathrm{N}(\mu, \sigma)$ måste vi beräkna integralen

$$\int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

[1] Vi definierar väntevärde och varians längre fram.

Detta kan man inte göra utan numeriska metoder. Man kan dock alltid återföra $\mathrm{N}(\mu, \sigma)$ till $\mathrm{N}(0, 1)$, den s.k. standardiserade normalfördelningen (Se nedan). Därmed kan man använda tabellen för $\mathrm{N}(0, 1)$ för *alla* normalfördelningar. Därför har man tabellerat värdena på integralen då $\mu = 0$ och $\sigma = 1$ (Se sidan **??**.), d.v.s. fördelningsfunktionen för $\mathrm{N}(0, 1)$, som vi kommer att beteckna Φ. Motsvarande frekvensfunktion betecknar vi med φ.

$$\Phi(b) \quad = \quad \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \tag{10}$$

I princip använder man *variabelsubstitution* för att visa övergången mellan $\mathrm{N}(\mu, \sigma)$ till $\mathrm{N}(0, 1)$:

$$\int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \left\{ \begin{array}{l} \frac{x-\mu}{\sigma} = t \Rightarrow dx = \sigma dt \\ \frac{a-\mu}{\sigma} = b \end{array} \right\} =$$

$$= \int_{-\infty}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} dt\sigma = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-t^2/2\sigma^2} dt \text{ där } b = \frac{a-\mu}{\sigma}$$

Detta betyder att med $a = x$ så är

$$F(x) = P(\xi \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{11}$$

där F är fördelningsfunktionen för ξ.

Ex 4. 9 En elkabels diameter (enhet: dm) är normalfördelad med parametrar $\mu = 0.8$ och $\sigma = 0.02$. Vad är sannolikheten att diametern

a) är högst 0.82 dm,

b) är mer än 0.81 dm,

c) är mellan 0.77 dm och 0.83 dm?

Lösning

Låt ξ vara elkabelns diameter. Då gäller $\xi \in \mathrm{N}(0.8, 0.02)$.

a) Vi söker $P(\xi \leq 0.82)$ och använder sambandet (11) och tabell för att beräkna den.

$$P(\xi \leq 0.82) = \Phi\left(\frac{0.82 - 0.8}{0.02}\right) = \Phi(1) = 0.8413 \approx 0.84$$

b)

$$P(\xi > 0.81) = 1 - P(\xi \leq 0.81) = 1 - \Phi\left(\frac{0.81 - 0.8}{0.02}\right) =$$

$$1 - \Phi(0.5) = 1 - 0.695 = 0.3085$$

c)

$$\begin{aligned}
P(0.77 < \xi \leq 0.83) &= P(\xi \leq 0.83) - P(\xi \leq 0.77) = \\
&= \Phi\left(\frac{0.83 - 0.8}{0.02}\right) - \Phi\left(\frac{0.77 - 0.8}{0.02}\right) = \\
&= \Phi(1.5) - \Phi(-1.5) = \\
&= \Phi(1.5) - (1 - \Phi(1.5)) = 2\Phi(1.5) - 1 = \\
&= 2 \cdot 0.9332 - 1 = 0.8664 \approx 0.87
\end{aligned}$$

■