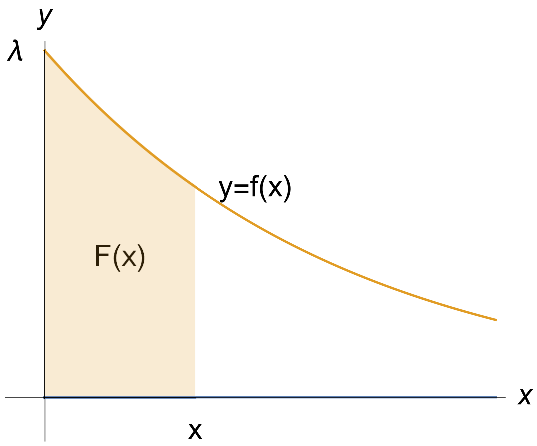
1 Kontinuerliga fördelningar

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

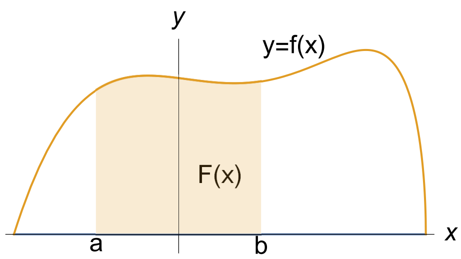
$$\int_{-\infty}^{\infty} f(t)dt = 1 \tag{1}$$

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

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



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

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

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

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

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 $\xi \in \text{Exp}(\lambda)$. 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

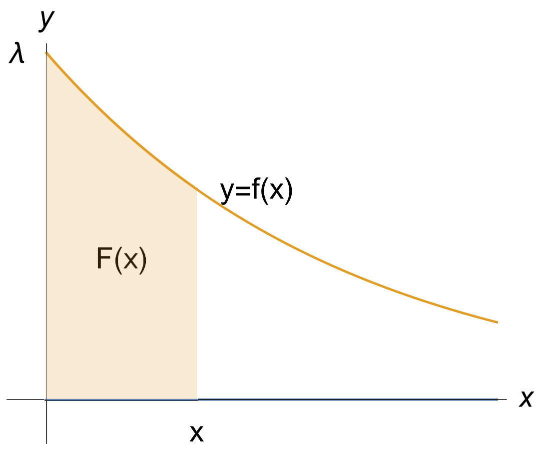
$$F(t) = \int_{-\infty}^t f(x)dx = \int_0^t \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^t = -(e^{-\lambda t} - 1) = 1 - e^{-\lambda t}$$

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 fördelingsfunktioner.

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 grått (**??**explivs).

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.

▪

Kommentarer

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 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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty, \tag{7}$$

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

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

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.

$$\Phi(b) \quad = \quad \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \tag{8}$$

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

$$F(x) = P(\xi \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{9}$$

där F är fördelningsfunktionen för ξ.

Ex 4. 5 En elkabels diameter (enhet: dm) är normalfördelad med parametrar $\mu = 0.8$ och $\sigma = 0.02$. Vad är sannolikheten att diametern

[1] Vi definierar väntevärde och varians längre fram.

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 (9) 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)

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

■

1.4 $t-$fördelning

En fördelning som liknar normalfördelningen och som vi kommer att använda längre fram är $t-$fördelningen. Till formen påminner dess frekvensfunktion om $\mathrm{N}(0,1)-$ fördelningens dito. Frekvensfunktionen för $t-$fördelningen med $n-1$ frihetsgrader är

$$f_{n-1}(x) := k_n \left(1 + \frac{x^2}{n-1}\right)^{-n/2}, \quad n = 2, 3, \ldots \tag{10}$$

där k_n är en normeringskonstant.

Figur 6: Normalfördelningen $\mathrm{N}(0,1)-$ (streckad) och t_3- fördelningen

Ex 4. 6 För $n = 2$ är alltså frekvensfunktionen $f_1(x) = \frac{k_2}{1+x^2}$. Denna fördelning kallas cauchyfördelningen. För att beräkna normeringskonstanten k_2 integrerar vi.

$$\int_{-\infty}^{\infty} \frac{k_2 dx}{1+x^2} = [\arctan x]_{-\infty}^{\infty} = k_2(\frac{\pi}{2} - (-\frac{\pi}{2})) = 1$$

d.v.s. $k_2 = 1/\pi$. Motsvarande fördelningsfunktion ges av

$$F(x) := \frac{1}{\pi}\int_{-\infty}^{x} \frac{dt}{1+t^2} = \ldots = \frac{\arctan x}{\pi} + \frac{1}{2}$$

Observera att $F(x) \to 1$ då $x \to \infty$ och att $F(x)$ växer (asympotiskt) mot 1.

Figur 7: Fördelningsfunktionen i **??**arctanfrd

■

Ex 4. 7 Beräkna k_4 (d.v.s. för den fördelning som har 3 frihetsgrader).

Lösning

Således är $n = 4$. Vi skall alltså beräkna integralen

$$k_4 \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{3}\right)^{-2} dx = 2k_4 \int_{0}^{\infty} \left(1 + \frac{x^2}{3}\right)^{-2} dx$$

Denna integral skall ju ha värdet 1. Vi gör substitutionen $x/\sqrt{3} = \tan t$. Då blir

$$\frac{dx}{\sqrt{3}} = (1 + \tan^2 t)dt$$

Insatt i den sista integralen får vi

$$2k_4 \int_{0}^{\infty} \left(1 + \frac{x^2}{3}\right)^{-2} dx = 2k_4\sqrt{3}\int_0^{\pi/2} \frac{1+\tan^2 t}{(1+\tan^2 t)^2} dt =$$

$$= 2k_4\sqrt{3}\int_0^{\pi/2} \frac{dt}{1+\tan^2 t} = 2k_4\sqrt{3}\int_0^{\pi/2} \cos^2 t dt =$$

$$= 2k_4\sqrt{3}\int_0^{\pi/2} \frac{1+\cos 2t}{2} dt = \ldots = 2k_4\sqrt{3}\cdot\frac{\pi}{4} = 1$$

Detta ger $k_4 = \frac{2}{\pi\sqrt{3}}$.

■

Vi återkommer till t−fördelningen i kapitel **??**.

1.5 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

$$F(x) \quad = \quad \begin{cases} 0, & x < 0 \\ \\ 1 - e^{-x^{\beta}/\alpha}, & x \geq 0 \end{cases} \tag{11}$$

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

1.6 $\Gamma-$fördelningen

En fördelning där frekvensfunktionen är

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{\lambda^n}{(n-1)!} \cdot x^{n-1} \cdot e^{-\lambda x} x \geq 0 \end{cases} \tag{12}$$

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. 8 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. 9 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. 10 Avgör vilka funktioner som är fördelningsfunktioner.

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

■

uppg. 1 Bestäm parametern λ för exponentialfördelningarna givna i figur 2 sidan 5.Bestäm motsvarande väntevärden.Beräkna sannolikheten $P(\xi \leq 3)$, om ξ har dessa frekvensfunktioner.

uppg. 2 Visa att en exponentialfördelad stokastisk variabel beskriver "icke-åldrande" (Jämför med **??**explivs), d.v.s.

$$P(\xi \geq x) = P(\xi \geq x + y | \xi \geq y) \text{ om } \xi \in \exp \lambda$$

uppg. 3

Två parallellkopplade ledningar i ett elektriskt system går via komponenterna A och B. Dessa komponenters hållbarhet är exponentialfördelade och oberoende med samma väntevärde 400 h. Systemet fungerar så länge minst en av komponenterna fungerar.

Vad är sannolikheten att systemet fungerar efter 300 h? Man har fem sådana oberoende system. Vad är sannolikheten att minst två av fem fungerar längre än 300 h?

uppg. 4 För normalfördelningen $\xi \in \mathrm{N}(0, 1)$ finns en tabell (sidan **??**). Beräkna med dess hjälp x_0, så att

$$P(\xi \leq x_0) = 0.90 P(-x_0 \leq \xi \leq x_0) = 0.95 P(-x_0 \leq \xi \leq x_0) = 0.3 P(x_0 \leq \xi) = 0.7$$

uppg. 5 Vikten på en fabriksproducerad brödsort är normalfördelad $\mathrm{N}(\mu, \sigma)$ (gram). Man vill att 99% av bröden skall väga minst 800 gram. Antag att $\sigma = 10$. Vilket är det minsta väntevärdet som man kan ha? Antag att man (högst) vill ha väntevärdet 810 gram. Vilken är den största standardavvikelsen som är möjlig?

uppg. 6 För en exponentialfördelning har man att $P(\xi \leq 400) = 0.7$. Beräkna parametern λ.

uppg. 7 Livslängden ξ för en elektronisk komponent är $\exp 0.002$ (h^{-1}). Beräkna (enhet h)

$$P(\xi > 400) P(\xi \leq 250) P(\xi > 400 | \xi > 250)$$

uppg. 8 En Geiger-Müller räknare registrerar radioaktiva sönderfall. Antalet registrerade sönderfall per minut är λ med $\lambda = 5.5$. Tiden mellan två (registrerade) sönderfall, kan man visa, är $\exp(\lambda)$ med samma λ. Beräkna sannolikheten att det är färre än 4 registrerade sönderfall under en given minut.Vilken är den förväntade tiden mellan två sönderfall?Beräkna sannolikheten att tiden mellan två sönderfall är < 10 sekunder.

uppg. 9 Livslängden ξ (i antal år) på en sorts utombordsmotorer kan beskrivas med en weibullfördelning med parametrar $\alpha = 7.0$ och $\beta = 1.2$. Vad är sannolikheten att motorn håller längre än 3 år?

uppg. 10 Resebyrån Bolivar arrangerar resor till La Paz. Resebyrån utgår från att restiden är $norm251.5$ (h). Den vill ge en "garantitid" t så att en flygresa med 95% sannolikhet inte tar längre tid än denna tid. Bestäm (den minsta) tiden t, som är möjlig.

uppg. 11 Tjockleken på en kopparkabel är 20.00.1 (mm). Vad är sannolikheten att den är tjockare än 20.05 mm? Vad är sannolikheten att den är tjockare än 20.05 mm, givet att den är smalare än 20.1 mm?Vad är sannolikheten att den är tjockare än 19.95 och tunnare än 20.05 mm?

2 Väntevärde, median och varians

2.1 Väntevärde

Vi skall nu *definiera* väntevärde för en kontinuerlig stokastisk variabel ξ. Vi utgår från definitionen av väntevärde för en diskret stokastisk variabel.Vi byter $P(\xi = x)$ mot en frekvensfunktion $f(x)$. Låt ξ vara en kontinuerlig stokastisk variabel som alltså antar värden i $\mathbb{R}$ och har frekvensfunktionen f. Nu approximerar vi den kontinuerliga stokastiska variabeln med en diskret stokastisk variabel, som antar värdena x_i, $i = 1, 2, \ldots, n$, med sannolikhet som ges i identitet. Den diskreta stokastiska variabeln har då väntevärdet

$$\sum_{x}^{n} x f(x).$$

Väntevärdet för den kontinuerliga stokastiska variabeln ξ och gör följande definition.

Definition 2
Låt ξ vara en kontinuerlig stokastisk variabel med frekvensfunktionen f. Då definieras väntevärdet för ξ som

$$E(\xi) = \int_{-\infty}^{\infty} x f(x) dx \tag{13}$$

2.2 Median

Medianen definieras som det $x-$värde, betecknat , sådant att

$$\int_{-\infty} f(x) dx = 1/2 \tag{14}$$

Kommentarer

i) Väntevärdet är tyngdpunkten för kurvans graf i $x-$led.

ii) För en symmetrisk fördelning sammanfaller väntevärdet (om det existerar) med medianen.

iii) Två stokastiska variabler ξ och η som är likafördelade, d.v.s. $P(\xi \leq x) = P(\eta \leq x)$ för alla x har samma väntevärde och varians.

Figur 8: Medianen delar ytan i två delar med 50% av arean på vardera sidan.

Ex 4. 11 Exponentialfördelningen har ju sannolikhets- eller frekvensfunktion

$$f(x) = \lambda e^{-\lambda x}, x \geq 0 \text{ och } f(x) = 0, x < 0, \text{ där } \lambda > 0$$

Väntevärdet fås då med partiell integration (P.I.)

$$\int_0^\infty x f(x) dx = \{\text{P.I.}\} = \left[-x e^{-\lambda x} - \frac{e^{-\lambda x}}{\lambda} \right]_0^\infty = \ldots = \frac{1}{\lambda}$$

Medianen ges av ekvationen

$$\int_0^{md} f(x) dx = \int_0 \lambda e^{-\lambda x} dx = 1/2$$

d.v.s.

$$e^{-\lambda} = \frac{1}{2} \Leftrightarrow \lambda = \ln 2 \Leftrightarrow = \frac{\ln 2}{\lambda}$$

Vi ser att v.v. $\mu = 1/\lambda$ och $= \ln 2/\lambda \approx 0.7/\lambda$. Speciellt är alltså $\mu >$.

■

Figur 9: Frekvensfunktionen $f(x) = 0.004 e^{-0.004x}$ för $x \geq 0$ från **??**explivs sidan 4. Obervera att väntevärdet $\mu = 1/\lambda = 1/0.004 = 250$.

2.3 Varians

I avsnitt **??** kom vi fram till att variansen $V(\xi) = E((\xi - \mu)^2)$ och standardavvikelsen $\sqrt{V(\xi)}$ är lämpliga spridningsmått för en sannolikhetsfördelning. För att beräkna dessa spridningsmått då ξ är en kontinuerlig stokastisk variabel så kan vi resonera på samma sätt som för väntevärdesdefinitionen. Vi får då följande definition.

Definition 3
Låt ξ vara en kontinuerlig stokastisk variabel med frekvensfunktion f och väntevärde $E(\xi) = \mu$. Variansen för ξ definieras som

$$V(\xi) = E((\xi - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \tag{15}$$

Standardavvikelsen för ξ definieras som $\sigma = \sqrt{V(\xi)}$.

På samma sätt som i det diskreta fallet gäller att om man ser sannolikhetsfördel- ningen som en massfördelning så svarar variansen mot tröghetsmomentet för denna massfördelning. Variansen $\sigma^2 = V(\xi)$ och standardavvikelsen σ för ξ är spridningsmått för sannolikhetsfördelningen till ξ. Ovanstående formel används sällan vid praktiska beräkningar av variansen. Istället görs följande omskrivning av definitionsformeln

$$\begin{aligned} V(\xi) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} (x^2 + \mu^2 - 2\mu x) f(x)dx = \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx + \mu^2 \underbrace{\int_{-\infty}^{\infty} f(x)dx}_{=1} - 2\mu \underbrace{\int_{-\infty}^{\infty} x f(x)dx}_{=\mu} = \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 \end{aligned}$$

Man kan visa att $\int_{-\infty}^{\infty} x^2 f(x)dx$ är väntevärdet för den stokastiska variabeln ξ^2 dvs $E(\xi^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$. Samma omskrivning som ovan kan göras om ξ är en diskret stokastisk variabel också. Den enda skillnaden i resonemanget är att alla integraltecken byts mot summatecken. Alltså gäller följande resultat.

Sats 2

$$V(\xi) = E(\xi^2) - \mu^2$$

$$\text{där } E(\xi^2) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x^2 f(x)dx & \text{om } \xi \text{ är kontinuerlig} \\ \displaystyle\sum_i x_i^2 P(\xi = x_i) & \text{om } \xi \text{ är diskret} \end{cases} \tag{16}$$

och $\mu = E(\xi)$

Ex 4. 12 Beräkna varians och standardavvikelse för väntetiden i **??**vantetid.

Lösning

Variansen beräknar vi med en integral. Väntevärdet är ju 5.

$$\int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = \int_0^{10} \frac{x^2 dx}{10} - 25 = \frac{100}{3} - 25 = \frac{25}{3}$$

Standardavvikelsen är alltså $\sigma = \dfrac{5}{\sqrt{3}}$.

■

Ex 4. 13 För att beräkna variansen för exponentialfördelningen börja vi med att beräkna integralen

$$\int_{\mathbb{R}_+} x^2 \lambda e^{-\lambda x} dx = \{\text{P.I.}\} = -\left[x^2 e^{-\lambda x}\right]_0^\infty + 2\int_0^\infty x e^{-\lambda x} dx$$

Den första utintegrerade termen $= 0$ och beräkningen av den återstående integralen är i princip gjord i **??**expvante sidan **??**, där väntevärdet beräknades till $\dfrac{1}{\lambda}$.

$$2\int_{\mathbb{R}_+} x e^{-\lambda x} dx = \frac{2}{\lambda}\int_{\mathbb{R}_+} x \lambda e^{-\lambda x} dx = \frac{2}{\lambda}\cdot\frac{1}{\lambda} = \frac{2}{\lambda^2}$$

Variansen och standardavvikelsen blir således

$$\sigma^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \text{ respektive } \sigma = \frac{1}{\lambda}$$

■

2.4 Kvantiler

Ex 4. 14 Betrakta $\xi \in 01$. Antag att vi vill finna det x sådant att $\Phi(x) \equiv P(\xi \leq x) = 0.95$. Vi kan då läsa “baklänges” i tabellen på sidan **??**. Efter lite letande finner vi värdet 1.645. Vi betecknar detta x med $k_{0.05}$, d.v.s. $x = k_{0.05} = 1.645$.

■

Med en kvantil k_α för en fördelning menas det “$x-$”värde sådant att

$$P(\xi \leq k_\alpha) = 1 - \alpha, \text{ d.v.s. } P(\xi > k_\alpha) = \alpha \tag{17}$$

Istället för beteckningen k_α använder man

i) λ_α för 01−fördelningen.

ii) $t_{n,\alpha}$ för $t-$fördelningen.

“Kvantil” k_α brukar stå för att arean t.v. om detta värde är α men i denna bok betecknar den alltså att arean t.h. är α.

uppg. 12 Visa att funktionera nedan är frekvensfunktioner. Beräkna också deras väntevärde och varians. $f(x)0x < 1\frac{3}{x^4}x \geq 1 f(x)\frac{2}{\pi}\cos^2 x - \pi/2 \leq x \leq \pi/20x < -\pi/2$ eller $x > \pi/2 f(x)0x <$

uppg. 13 För en normalfördelad stokastisk variabel ξ vet man att $P(\xi \leq 1) = 0.05$ och $P(\xi > 10) = 0.025$. Bestäm μ och σ för normalfördelningen.

uppg. 14 Beräkna variansen för $\xi \in \mu\sigma$, d.v.s. visa att dess varians är σ^2.

uppg. 15 Beräkna varians och standardavvikelse i **??**feltol sidan 2.

Figur 10: $k_{0.05}-$kvantil

Figur 11: Kvantilen $x = k_{0.05}$ för standardnormalfördelningen som har en egen beteckning $\lambda_{0.05}$.

uppg. 16
Antag att ξ har frekvensfunktionen

$$f(x)0x < 0\frac{2x}{(1+x^2)^2}x \geq 0$$

Beräkna medianen, maximipunkten och väntevärdet. Antag att ξ är $\Gamma(2, \lambda)-$fördelad. Beräkna frekvensfunktionens väntevärde och dess maximipunkt.

uppg. 17 En stokastisk variabel ξ har frekvensfunktionen

$$f(x) = \begin{cases} 0, & x < 1 \\ \\ \dfrac{a}{x^2}, & x \geq 1 \end{cases}$$

a) Bestäm konstanten a så att funktionen blir en frekvensfunktion.

b) Bestäm $P(0 < \xi \leq 3)$.

c) Bestäm funktionens median.

d) Visa att funktionen saknar väntevärde.

uppg. 18 En stokastisk variabel ξ har följande fördelningsfunktion

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x^2/2}, & x \geq 0 \end{cases}$$

a) Vad är sannolikheten att $\xi > 1.5$?

b) Vad är väntevärdet för ξ?

c) Beräkna variansen.

uppg. 19 Ta fram följande kvantiler m.h.a. av tabellerna på sidorna **??** och **??**.

$$\lambda_{0.01}\lambda_{0.10}\lambda_{0.025}t_{10,0.01}t_{19,0.10}t_{29,0.025}$$

uppg. 20 Betrakta funktionen

$$f(x) = \frac{A}{e^x + e^{-x}}, \quad x \in \mathbb{R}$$

Bestäm konstanten A så att f blir en frekvensfunktion. Vad blir dess väntevärde?Beräkna dess kvantil $k_{0.10}$.* Man kan beräkna dess standardavvikelse, som blir $\frac{\pi}{2}$. Kan du visa det?