

PRO GRADU -TUTKIELMA

Olli Iltanen

Luotettavuusanalyysin perusteita

TAMPEREEN YLIOPISTO
Matematiikan, tilastotieteen ja filosofian laitos
Tilastotiede
Huhtikuu 2008

Tampereen yliopisto

Matematiikan, tilastotieteen ja filosofian laitos

ILTANEN, OLLI: Luotettavuusanalyysin perusteita

Pro gradu -tutkielma, 39 s., 8 liites.

Tilastotiede

Huhtikuu 2008

Tiivistelmä

Tässä tutkielmassa johdatellaan lukijaa luotettavuusanalyysin perusteisiin kolmen tunnetun jakauman ja esimerkkien avulla. Aluksi käydään läpi sellaisia luotettavuuden käsitteitä kuin luotettavuusfunktio, hasardifunktio ja vikaantumistaajuus sekä tutustutaan datatyyppeihin, joita luotettavuusanalyysissä kohdataan. Tunnetuista elinaikajakaumista on mukaan otettu eksponenttijakauma ja lognormaalijakauma. Eksponenttijakauman vikaantumistaajuus on ajasta riippumaton, minkä ansiosta jakaumalla on erityisasema luotettavuusanalyysissä. Eksponenttijakaumalle on omistettu oma lukunsa, jossa käsitellään muun muassa unohtamisominaisuutta ja jakauman parametrin estimointia sekä koesuunnittelua. Lognormaalijakautuneista vikaantumisajoista saadaan yksinkertaisella muunnoksella normaalijakautuneita, mistä johtuen myös normaalijakauma kuuluu tutkielman aihepiiriin. Normaalijakauman käsittelyssä on parametrien estimointi keskeisellä sijalla, mutta myös joitakin jakauman perusominaisuuksia kerrataan. Tässä yhteydessä esitellään myös esimerkkiaineisto. Lognormaalijakauma käydään läpi lyhyemmin, sillä analyysissä voidaan käyttää normaalijakauman menetelmiä. Viidennessä luvussa käsitellään kahta luotettavuusanalyysille tärkeää aihetta: nopeutettua vikaantumista ja Coxin verrannollisten hasardien regressiomallia. Nopeutetun vikaantumisen malleista on esimerkiksi valittu Arrheniuksen malli.

Sisältö

1	Johdanto	7
2	Luotettavuuden käsitteitä	9
2.1	Luotettavuusfunktio	9
2.2	Hasardifunktio ja kumulatiivinen hasardifunktio	9
2.3	Keskimääräinen vikaantumistaajuus	11
2.4	Mittayksiköitä	11
2.5	Hasardifunktion kuvaaja	11
2.6	Datan tyypit	12
2.6.1	Ensimmäisen tyypin sensuroitu data	12
2.6.2	Toisen tyypin sensuroitu data	12
2.6.3	Ryhmitelty data	12
2.6.4	Monisensuroitu data	13
3	EkspONENTTijakauma	14
3.1	EkspONENTTijakauman perusteita	14
3.2	MTTF	15
3.3	Unohtamisominaisuus	15
3.4	Lambdan estimoinnista	16
3.5	Lambdan ja MTTF:n luottamusvälit	17
3.6	Koesuunnittelu eksponenttijakauman tapauksessa	17
3.7	Simuloitu esimerkkitapaus	18
3.7.1	Satunnaisotos eksponenttijakaumasta	18
3.7.2	χ^2 -yhteensopivuustesti	18
3.7.3	Esimerkki	20
4	Normaalijakauma ja lognormaalijakauma	22
4.1	Normaalijakauman perusteita	22
4.2	Keskeinen rajaväittäjä	23
4.3	Normaalijakauman parametrien estimointi	24
4.4	Luottamusväli μ :lle ja σ :lle	26
4.5	Lognormaalijakauma	26
4.6	Lognormaalijakauman parametrien estimointi	27
4.7	Esimerkki lognormaalijakautuneella datalla	28
4.7.1	Paukkumaissiaineisto	28
4.7.2	Sensuroimattoman datan parametrien estimaatit	28

5	Coxin regressio ja nopeutettu vikaantuminen	30
5.1	Verrannollisten hasardien malli	30
5.2	Mallin parametrien estimoinnista	31
5.3	Muuttujien valinnasta	32
5.4	Kaksi esimerkkiä Coxin regressioanalyysistä	33
5.4.1	Alkuperäinen paukkumaissiaineisto	33
5.4.2	Pieni paukkumaissiaineisto	34
5.5	Nopeutuskerroin	35
5.6	Arrheniuksen malli	36
5.7	Nopeutuskerroin paukkumaissiaineistolle	37
6	Johtopäätökset	38
	Lähdeluettelo	40
	Liite 1: R-koodi eksponenttijakaumaesimerkille	41
	Liite 2: R-koodi lognormaalijakaumaesimerkille	44
	Liite 3: Coxin regression kuvaajia	46
	Liite 4: R-koodi nopeutuskertoimen laskemiseksi	48

1 Johdanto

Niin kauan kuin insinöörit ovat suunnitelleet ja kehittäneet laitteita ja niiden osia, on luotettavuuden parantaminen ollut osa kehitysprosessia. Tätä yrityksen ja erehdyksen periaatetta on sovellettu jo kauan ennen kuin järjestelmällinen datan kerääminen ja analysointi alkoi, koska laitteen rikkoontumista tai vikaantumista ei yleensä voi olla huomaamatta. Mikäli tuote ei kestä käytössä, on tuotekehitystä jatkettava.

Vielä 1900-luvun alussa suunnittelijoita ei ahdistettu jatkuvalla kiireellä ja huolilla tuotannon kustannustehokkuudesta siinä määrin kuin tänä päivänä. Luotettavuusanalyysin tekniikoiden tarvetta laitteiden suunnittelu- ja kehitysvaiheessa ei tunnustettu tai sitä ei ollut olemassa. Näin ollen myöskään vikaantumistietoja ei järjestelmällisesti kerätty. Myös 1900-luvun alun tuotantotavoilla oli syynä luotettavuusanalyysin tarpeen vähäisyyteen, laitteiden osat valmistettiin vielä suurimmaksi osaksi perinteisin käsityömenetelmin verstaasympäristössä, jossa ammattimies vastasi työnsä laadusta. Luotettavuuden käsitteen varhaiset perustukset löytyvätkin tuotantoteollisuuden sijaan vakuutusalan alalta, erityisesti ihmisten eliniänodotteen tutkimuksesta.

Jotta luotettavuutta voitaisiin mitata riittävällä tarkkuudella, on luotettavuuden määrittelyn oltava täsmällinen: Luotettavuus on todennäköisyys, että laitteisto, laite tai laitteen osa suoriutuu tehtävästään yli annetun ajan, olosuhteissa, jotka on ennalta määrätty (Ebeling 1997). Se on siis todennäköisyys, että laitteeseen ei tule ajan myötä vikaa.

Aina ei ole helppoa päästä yhteisymmärrykseen siitä, milloin laite suoriutuu tehtävästään eli toimii oikein. Tämä luotettavuuden osatekijä on suurin yksittäinen kiistan aiheuttaja laitteen myyjän ja ostajan välillä (Leitch 1995). Onkin ehdottoman tärkeää määritellä etukäteen kaikki ne toimintahäiriöt ja poikkeavuudet, joita voidaan pitää vikoina. Määrittelyssä voidaan hyödyntää periaatetta, jonka mukaan viaksi voidaan lukea mikä hyvänsä laitteen puute, joka aiheuttaa ennenaikaisen huollon tarpeen tai muutoin estää laitteen normaalin käytön (Leitch 1995). Yleensä vikaantuminen aiheutuu joko laitteen äkillisestä hajoamisesta tai ikääntymisen aiheuttamasta kulumisesta tai materiaalien väsymisestä.

Minkä hyvänsä laitteen luotettavuus riippuu voimakkaasti ympäröivistä olosuhteista. Tyypillisiä luotettavuuteen vaikuttavia ympäristötekijöitä ovat muun muassa ilmasto-, pakkaus-, kuljetus-, varastointi- ja asennusolosuhteet, käyttäjä, huollon saatavuus, pöly sekä ilmansaasteet (Leitch 1995). Käyttöympäristön tutkiminen on olennainen osa luotettavuuden estimointia.

Luotettavuus vähenee ajan myötä, sillä mitä kauemmin laite on käytössä sitä suurempi on vikaantumisen todennäköisyys. Käyttöikää ei aina mitata aikayksiköillä, vaan se saattaa muodostua myös esimerkiksi kilometreistä, jotka ajoneuvo on taittanut tai sykleistä (Ebeling 1997). Sykleinä voidaan laskea vaikka lentokoneen laskeutumisten lukumäärä tai sähkömoottorin käynnistysten lukumäärä.

Sarjatuotannon myötä syntyi tarve komponenttien tasalaatuisuuden varmistamiseen, standardisointiin. Vialliset osat pyrittiin tunnistamaan jo tuotantovaiheessa tarkistuksin ja kokein, ja luotettavuutta alettiin kontrolloida laadunvalvontaa tehostamalla. Sähkövoiman käytön yleistyminen, jota toinen maailmansota kiihdytti, johti yhä monimutkaisempien komponenttien sarjatuotannon tarpeeseen. Asevoimien laitteiden jatkuva epäluotettavuus kenttäolosuhteissa 1940- ja 1950-luvuilla kohdisti huomion parempien luotettavuusanalyysin menetelmien tarpeeseen. Tietoa vikaantumisista alettiin kerätä laajemmassa mittakaavassa sekä laitteiden luonnollisissa käyttöoloissa että laboratoriotestien yhteydessä.

Ruotsalainen insinööri, tiedemies ja matemaatikko Waloddi Weibull (1887–1979) tutki 1930-luvun lopulla ja 1940-luvulla materiaalien lujuutta ja väsymistä ja johti tutkimuksistaan tiheysjakauman, joka kantaa hänen nimeään (Ebeling 1997). Weibull-jakauma on eksponenttijakauman yleistys ja sopii joustavuutensa ansiosta monenlaisten vikaantumismekanismien mallintamiseen (Tobias & Trindade 1986). 1950-luvulla julkaistuista luotettavuuden alan tutkimuksista valtaosa keskittyi vielä eksponenttijakauman sovelluksiin, mutta vuosikymmenen lopulla Weibull-jakauman merkitys lisääntyi (Ebeling 1997).

Lognormaalijakauma muistuttaa ominaisuuksiltaan Weibull-jakaumaa. Myös lognormaalijakaumaa käytetään yleisesti luotettavuusanalyysissä ja sen houkuttelevuutta lisää läheinen yhteys normaalijakaumaan. Tässä tutkielmassa tarkastellaan lähemmin eksponentti-, normaali- ja lognormaalijakaumia. Liikkeelle lähdetään luotettavuuden peruskäsitteiden esittelyllä, jonka jälkeen käsitellään luotettavuus- tai elinaikajakaumista yksinkertaisinta, eksponenttijakaumaa. Yksinkertaisen esimerkin jälkeen kerrataan joitakin normaalijakauman ominaisuuksia ja esitellään parametrien estimointiprosessi sekä sensuroimattoman että sensuroidun datan tapauksissa. Tämän jälkeen tutustutaan lognormaalijakaumaan sekä lognormaalijakauman ja normaalijakauman yhteyteen. Lopuksi käydään vielä läpi joitakin nopeutettuun vikaantumiseen ja Coxin regressioon liittyviä aiheita ja esimerkkejä.

Vaikka runsas joukko luotettavuusanalyysin kannalta keskeisiä aiheita rajautuu tutkielman ulkopuolelle, toimii tämä teksti toivottavasti lyhyenä johdantona aiheeseen. Mukaan on valittu kokonaisuuksia, jotka useimmille tilastotieteen opiskelijoille ovat jo entuudestaan tuttuja ja siksi helposti lähestyttäviä.

2 Luotettavuuden käsitteitä

2.1 Luotettavuusfunktio

Luotettavuus on suure, joka kuvaa kuinka harvoin vikoja ilmenee ajan kuluessa. Laitteet tai useamman laitteen järjestelmät muodostavat populaatioita. Laitteiden elinikää voidaan kuvata kyseisten populaatioiden teoreettisilla malleilla. Kertymäfunktio on tällainen malli, joka määrittää populaation elinaikajakautuksen. Käytetään kertymäfunktioista merkintää $F(t)$, jossa t viittaa aikaan. $F(t)$ voidaan tulkita todennäköisyydeksi, että populaatiosta satunnaisesti poimittu alkiio lakkaa toimimasta aikaan t mennessä. Toisaalta $F(t)$ voidaan ymmärtää populaation niiden alkioiden osuudeksi, jotka lakkaavat toimimasta aikaan t mennessä.

Usein ollaan vikaantuneiden alkioiden sijaan kiinnostuneita yhä toimivista alkiioista. Määritellään luotettavuusfunktio

$$R(t) = 1 - F(t).$$

Kuten kertymäfunktioilla, myös luotettavuusfunktioilla on kaksi hyödyllistä tulkintaa: $R(t)$ voidaan ajatella todennäköisyydeksi, että populaatiosta satunnaisesti poimittu alkiio toimii yhä ajan t jälkeen, tai $R(t)$ on populaation niiden alkioiden osuus, jotka toimivat vähintään aikaan t saakka.

Jos kertymäfunktio $F(t)$ kuvaa populaatiota, josta poimitaan n kappaletta identtisiä alkiioita, niin $nF(t)$ on odotusarvo niiden alkioiden lukumäärälle, jotka ovat lakanneet toimimasta aikaan t mennessä. Vastaavasti, $nR(t)$ on odotusarvo niiden alkioiden lukumäärälle, jotka yhä toimivat ajan t jälkeen.

2.2 Hasardifunktio ja kumulatiivinen hasardifunktio

Laitteen vikaantumisen saattaa aiheuttaa yksittäinen fyysinen tekijä tai usean fyysisen tekijän yhteisvaikutus. Kaikkien fyysisten tekijöiden vaikutusta laitteeseen ei pystytä mallintamaan matemaattisesti, koska tekijöitä voi olla hyvin suuri määrä eikä kaikkia yhteisvaikutuksia tunneta. Siksi turvaututaan yksinkertaistettuun kuvaukseen ja yleensä mallina käytetään jotakin tunnettua tiheysjakaumaa. Sopivan jakauman löytämistä helpottaa hasardifunktio eli vikaantumistaajuus (hazard function, hazard rate, failure rate). Vakuutus-tieteessä hasardifunktio tunnetaan kuolleisuusasteena tai vaaran asteena (force of mortality), ääriarvoteoriassa sitä kutsutaan intensiteettifunktioksi (intensi-

ty function) ja taloustieteissä sen käänteislukua nimitetään Millin suhteeksi (Mill's ratio) (Mann, Schafer & Singpurwalla 1974).

Olkoon $P(A)$ todennäköisyys, että tapahtuma A sattuu ja $P(B | A)$ todennäköisyys, että tapahtuma B sattuu, kun A :n tiedetään sattuneen. Tällöin $P(B | A)$ voidaan laskea osamääränä

$$P(B | A) = \frac{P(A \cap B)}{P(A)}.$$

Ehdollisen todennäköisyyden kaavaa käyttämällä voidaan laskea todennäköisyys, että yksikkö lakkaa toimimasta lyhyessä ajassa Δt , toimittuaan ensin aikaan t saakka. Olkoon $A =$ 'yksikkö toiminut aikaan t saakka' ja $B =$ 'yksikkö lakkaa toimimasta ajassa Δt '. Tällöin

$$P(B | A) = \frac{F(t + \Delta t) - F(t)}{R(t)}.$$

Muutetaan osamäärä suhdeluvuksi jakamalla se Δt :llä, jolloin saadaan

$$\frac{F(t + \Delta t) - F(t)}{R(t)\Delta t}.$$

Annetaan Δt :n lähestyä nollaa, jolloin erotusosamäärän raja-arvo

$$\lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}$$

on kertymäfunktion derivaatta $F'(t)$. Koska $F'(t) = f(t)$, päädytään hasardifunktion yhtälöön

$$h(t) = \frac{f(t)}{R(t)}.$$

Hasardifunktio ilmoittaa vikaantuneiden yksiköiden lukumäärän aikayksikköä kohti. Se ei ole todennäköisyys ja voi siksi saada ykköstä suurempia arvoja, mutta on kuitenkin aina ei-negatiivinen. Toisinaan hasardifunktiona käytetään $f(t)$:tä, joka ei ota huomioon ennen ajankohtaa t vikaantuneiden vaikutusta (Tobias & Trindade 1986). Integroimalla hasardifunktio saadaan kumulatiivinen hasardifunktio $H(t)$,

$$H(t) = \int_0^t h(t)dt.$$

Integraalille löytyy ratkaisu

$$H(t) = -\ln R(t),$$

joka voidaan edelleen saattaa muotoon

$$F(t) = 1 - e^{-H(t)} = 1 - e^{-\int_0^t h(t)dt}.$$

Kun siis yksi funktioista $F(t)$, $f(t)$, $H(t)$ tai $h(t)$ tunnetaan, saadaan muut ratkaistua.

2.3 Keskimääräinen vikaantumistaajuus

Vikaantumistaajuus vaihtelee ajan myötä. Joissakin tilanteissa on hyödyllistä määrittää yksi arvo, joka kuvaa hasardifunktion käytöstä tietyllä aikavälillä (t_1, t_2) . Keskimääräinen vikaantumistaajuus (average failure rate, AFR) välillä (t_1, t_2) saadaan integroimalla hasardifunktio yli aikavälin ja jakamalla integraalin arvo erotuksella $t_2 - t_1$.

$$\begin{aligned} \text{AFR}(t_1, t_2) &= \frac{\int_{t_1}^{t_2} h(t) dt}{t_2 - t_1} \\ &= \frac{H(t_2) - H(t_1)}{t_2 - t_1} \\ &= \frac{\ln R(t_1) - \ln R(t_2)}{t_2 - t_1} \end{aligned}$$

Jos aikaväli on $(0, T)$, on AFR yksinkertaisesti

$$\text{AFR}(T) = \frac{H(T)}{T} = \frac{-\ln R(T)}{T},$$

joka on likimain $F(T)/T$ pienillä $F(T)$:n arvoilla (Tobias & Trindade 1986).

2.4 Mittayksiköitä

Laitteiden vikaantumistaajuudet ovat usein niin pieniä, että tunnissa vikaantuneiden lukumäärä on yksikkönä liian karkealla tasolla. Usein käytettyjä vikaantumistaajuuden yksiköitä ovat prosenttia tuhatta tuntia kohden (%/K) ja miljoonasosaa tuhatta tuntia kohden (PPM/K tai fails in time, FIT). Yhden prosentin vikaantumistaajuus tuhatta tuntia kohden tarkoittaa, että sadan komponentin toimiessa tuhannen tunnin ajan, on odotettavissa, että yksi niistä vikaantuu. Vastaavasti yhden miljoonasosan vikaantumistaajuus tuhatta tuntia kohden tarkoittaa, että miljoonan komponentin toimiessa tuhannen tunnin ajan, on odotettavissa, että yksi niistä vikaantuu.

2.5 Hasardifunktion kuvaaja

Hasardifunktio voi olla mikä hyvänsä ei-negatiivinen funktio $h(t)$, jonka integraali $H(t)$ lähenee ääretöntä, ajan lähetessä ääretöntä. Useimpien laitteiden hasardifunktion kuvaaja on käytännössä kuitenkin joko monotonisesti nouseva käyrä, monotonisesti laskeva käyrä tai U:n muotoinen käyrä, joka tunnetaan myös ammekäyränä. Kylpyammeen muotoinen kuvaaja syntyy etenkin tilanteissa, joissa populaation alkioita seurataan elinkaaren alusta loppuun. Hasardifunktion kuvaaja ihmispopulaatioille on tavallisesti ammeen muotoinen: Kuoleman riski välittömästi syntymän jälkeen ja varhaislapsuudessa on suuri. Tämän jälkeen seuraa pitkähkö tasainen alhaisen riskin kausi, jonka jälkeen riski alkaa kasvaa iän myötä noin kolmenkymmenen ikävuoden paikkeilla.

Nouseva käyrä hasardifunktion kuvaajana on tavallinen. Kiinnostavin vaihe laitteiden elämässä on usein silloin, kun ikääntymisen myötä ilmenevät viat alkavat lisääntyä. Toisinaan populaatioista, joiden hasardifunktio on kylpyammeen muotoinen, karsitaan varhaisessa vaiheessa vikaantuneet alkiot. Tällöin jäljelle jäävien alkioiden hasardifunktion kuvaaja on nouseva. Näin toimitaan esimerkiksi, kun tuotantoprosessi käsittää sisäänaajovaiheen. Sisäänaajovaiheessa tuotteita käytetään lyhyesti ennen kuin ne toimitetaan myyntiin. Näin populaatiosta saadaan karsittua tuotteet, jotka olisivat vikaantuneet elinkaarensa alussa.

On olemassa tärkeitä ja rakenteeltaan yksinkertaisia malleja, joiden hasardifunktion arvo on vakio. Laskeva käyrä hasardifunktion kuvaajana on harvinaisen, mutta mahdollinen.

2.6 Datan tyypit

2.6.1 Ensimmäisen tyypin sensuroitu data

Tehdään koe, jossa n laitetta käynnistetään yhtä aikaa. Kokeen suunniteltu kesto on T . Oletetaan, että laitteista r kappaletta vikaantuu, ja tarkat vikaantumisaikat ovat $t_1 \leq t_2 \leq t_3 \leq \dots \leq t_r \leq T$. Kokeen lopussa laitteista $n - r$ kappaletta toimii. Yhä toimivista laitteista tiedetään vain, että niiden elinikä on suurempi kuin T .

Tällaisesta kokeesta saatua dataa kutsutaan ensimmäisen tyypin sensuroiduksi dataksi (exact times — censored type I). Etuna on aikataulun pitävyys, koska kokeen kesto T on päätetty etukäteen. Koetyypin pahin ongelma on $r:n$ satunnaisuus: etukäteen ei tiedetä kuinka moni laitteista vikaantuu aikaan T mennessä. Vikaantumistaajuuden estimaatin tarkkuus riippuu vikaantuneiden laitteiden lukumäärästä r , ei otoskoosta n . Liian pieni otoskoko tai kokeen liian vähäinen kuormittavuus johtavat päätöksenteon kannalta riittämättömään tietoon.

2.6.2 Toisen tyypin sensuroitu data

Toisen tyypin sensuroitua dataa (exact times — censored type II) saadaan koeasetelmasta, joka on kuten ensimmäisen tyypin sensuroidun datan tapauksessa, mutta koetta ei lopeteta ennalta määrätyn ajan T kuluttua, vaan odotetaan kunnes täsmälleen r laitetta on vikaantunut ja lopetetaan. Koeasetelman tuottama tietämys vikaantumistaajuudesta tunnetaan etukäteen, koska r on kiinnitetty jo ennen kokeen alkua. Koeaika T on nyt satunnainen, ja aikataulut saattavat muodostua ongelmaksi.

2.6.3 Ryhmitelty data

Koeasetelma sekä ensimmäisen että toisen tyypin sensuroidun datan tapauksessa edellyttää välineitä tarkkojen vikaantumisaikojen mittaamiseen ja tal-

lentamiseen. Tällainen koeasetelma on usein epäkäytännöllinen. Tarkkojen vikaantumisaikojen tallentamisen sijasta voidaan määrittää ennalta ajankohdat $T_1, T_2, T_3, \dots, T_k$, jolloin tarkistetaan kuinka moni yksikkö on lakannut toimimasta. Tätä koemenettelyä voidaan kutsua datan ryhmittelyksi (readout time data, interval data, grouped data).

Kokeen alussa mukana on n yksikköä, joiden kestävyyttä testataan. Hetkellä T_1 kaikki yksiköt tarkistetaan ja vikaantuneet poistetaan. Vikaantuneiden lukumäärä hetkellä T_1 on r_1 . Koe jatkuu niiden $n - r_1$:n yksikön osalta, jotka yhä toimivat. Hetkellä T_2 tarkistetaan jälleen kaikki yksiköt, jotka ovat mukana kokeessa. Välillä $T_2 - T_1$ vikaantuneet r_2 yksikköä poistetaan, ja $n - r_1 - r_2$ yksikköä jatkaa koetta. Näin koetta jatketaan tarkistamalla yksiköiden toimivuus ajan hetkillä T_3, T_4, \dots, T_k . Viimeinen tarkistus on kokeen lopussa, hetkellä $T_k = T$. Etukäteen ei tiedetä kuinka moni yksikkö lakkaa toimimasta tietyllä aikavälillä. Tarkkoja vikaantumisaikoja ei tunneta.

Ryhmitellyn datan tapauksessa kohdataan samankaltaisia ongelmia kuin ensimmäisen tyypin sensuroidun datan yhteydessä. Koe saattaa päättyä ennen kuin riittävän moni yksikkö on lakannut toimimasta. Myös tarkkuus kärsii, kun täsmällisiä vikaantumisaikoja ei tunneta. Vaikka vikaantumisia olisi paljon, saattaa päättely olla mahdotonta, jos vikaantumiset ovat keskittyneet vain harvoille aikaväleille (Tobias & Trindade 1986).

2.6.4 Monisensuroitu data

Yleisimmässä tapauksessa kustakin koeyksiköstä saatetaan tuntea aikaväli, jolloin yksikön tiedetään vikaantuneen, tai aikaraja t_i , jonka yli yksikön tiedetään selvinneen. Kullakin yksiköllä saattaa olla oma aikavälinsä tai aikarajansa, joka poikkeaa kaikkien muiden yksiköiden aikaväleistä tai -rajoista. Kutsutaan tällaista dataa monisensuroiduksi (multicensored data). Päättelyn kannalta monisensuroitu data on datatyypeistä haastavin.

3 Eksponenttijakauma

3.1 Eksponenttijakauman perusteita

Eksponenttijakauman merkitys luotettavuustutkimuksissa on verrattavissa normaali-jakauman keskeiseen asemaan muilla tilastotieteen osa-alueilla (Mann et al. 1974). On monia tilanteita, joissa laitteen tai organismin elinaikaa ei voida kuvata eksponenttijakauman avulla. Siitä huolimatta luotettavuusanalyysin teorian tunteminen eksponenttijakauman tapauksessa helpottaa monimutkaisempien tilanteiden ymmärtämistä. Eksponenttijakauman hyödyllisyys perustuu sen yksinkertaisuuteen ja sen läheiseen yhteyteen Poissonin prosessin kanssa. Eksponenttijakauma soveltuu tilanteisiin, joissa vikaantumiset ilmenevät täysin satunnaisesti. Tämä voidaan tulkita siten, että olipa vikaantumisten syy mikä hyvänsä, ne ilmenevät Poissonin prosessin periaatteiden mukaisesti. Eksponenttijakauman sovellusalueita rajoittaa jakauman unohtamisominaisuus, ominaisuus, jota muilla jakaumilla ei ole (Mann et al. 1974).

Eksponenttijakauman tiheysfunktio on

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

ja kertymäfunktio

$$F(t) = 1 - e^{-\lambda t}.$$

Sekä eksponenttijakauman tiheys- että kertymäfunktiossa, on λ ainoa tunnetun parametrisen $F(t)$:n arvot voidaan helposti laskea millä hyvänsä t :n arvolla, kun λ tunnetaan.

Luotettavuusfunktio on

$$R(t) = 1 - F(t) = e^{-\lambda t},$$

joten hasardifunktio on

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Eksponenttijakauman vikaantumistaajuus on siis ajasta riippumaton vakio λ , joka kertoo vikaantumisten lukumäärän aikayksikköä kohti. Eksponenttijakauma on ainoa jakauma, jonka hasardifunktio on vakio (Tobias & Trindade 1986).

Keskimääräinen vikaantumistaajuus aikavälillä (t_1, t_2) saadaan osamäärästä

$$AFR = \frac{\ln R(t_1) - \ln R(t_2)}{t_2 - t_1}.$$

Eksponttijakauman tapauksessa osamäärä saadaan muotoon

$$\frac{\lambda(t_2 - t_1)}{(t_2 - t_1)} = \lambda.$$

3.2 MTTF

Elinaikajakauman keskiarvo on uuden yksikön odotettu elinaika eli odotettu aika yksikön vikaantumiseen. Tästä keskiarvosta käytetään lyhennettä MTTF (Mean Time To Fail). Eksponttijakaumalle on

$$MTTF = \int_0^{\infty} t\lambda e^{-\lambda t} dt.$$

Osittaisintegroimalla yhtälön oikea puoli päädytään muotoon

$$MTTF = \frac{1}{\lambda}.$$

Jos populaation vikaantumistaajuus on vakio, on MTTF tämän vakion käänteisluku.

Populaation sijaintia kuvaa myös vikaantumisaikojen mediaani, T_{50} . Mediaani on piste, jossa kertymäfunktio saavuttaa arvon $1/2$. Eksponttijakauman mediaani saadaan lausekkeesta

$$F(T_{50}) = 0.5 = 1 - e^{-\lambda T_{50}}.$$

Otetaan yhtälön molemmista puolista luonnollinen logaritmi ja ratkaistaan T_{50} :n suhteen, jolloin saadaan

$$T_{50} = \frac{\ln 2}{\lambda} \approx \frac{0.693}{\lambda}.$$

T_{50} on pienempi kuin MTTF, sillä osoittaja 0.693 on pienempi kuin 1.

$$F(MTTF) = 1 - e^{-\lambda/\lambda} = 1 - e^{-1} \approx 0.632,$$

joten likimain 63.2 % eksponttijakauman alkioista on vikaantunut aikaan $MTTF = 1/\lambda$ mennessä, kun vikaantumistaajuus on λ .

3.3 Unohtamisominaisuus

Eksponttijakauman alkioden vikaantumistodennäköisyys on riippumaton siitä, kuinka kauan alkiot ovat olleet toiminnassa. Jos $T \sim Exp(\lambda)$, niin

$$P(T > a + b | T > a) = P(T > b), \quad \forall a, b \geq 0.$$

Tuloksen todistamiseksi lasketaan ehdollinen todennäköisyys (Liski 2005, s. 162)

$$\begin{aligned}
 P(T > a + b | T > a) &= \frac{P(T > a \cap T > a + b)}{P(T > a)} \\
 &= \frac{P(T > a + b)}{P(T > a)} \\
 &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} \\
 &= e^{-\lambda b} \\
 &= P(T > b)
 \end{aligned}$$

Unohtamisominaisuudella on suuri merkitys koesuunnittelussa. Kymmenen yksikön koekäyttäminen 20000 tunnin ajan tuottaa saman informaation kuin tuhannen yksikön koekäyttäminen 200 tunnin ajan. Jos yksikkö vikaantuu kokeen aikana, se voidaan korjata tai korvata koeasetelmaa muuttamatta. Eri ikäiset yksiköt käyttäytyvät kuin ne olisivat saman ikäisiä.

3.4 Lambdan estimoinnista

Eksponenttijakauman tapauksessa estimoidaan vain yksi parametri, λ . Paras estimaatti λ :lle on

$$(3.1) \quad \hat{\lambda} = \frac{\text{vikaantuneiden lukumäärä}}{\text{kokeen kokonaiskesto}}.$$

Nimittäjä saadaan laskemalla yhteen kaikkien kokeeseen osallistuneiden laitteiden käyttötunnit.

Jos kaikki koeyksiköt lakkaavat toimimasta kokeen aikana ja vikaantumisten tarkat ajat tunnetaan, on λ :n estimaatti yksinkertaisesti otoskeskiarvon käänteisluku, $\hat{\lambda} = 1/(\text{otoksen MTTF})$.

Jos kokeessa on n yksikköä ja näistä r vikaantuu, niin ensimmäisen tyypin sensuroidulle datalle

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^r t_i + (n-r)T}.$$

T on kokeen ennalta valittu päättymisajankohta ja t_i :t ovat tarkkoja vikaantumisaikoja niille yksiköille, jotka lakkaavat toimimasta ennen kokeen päättymistä.

Jos koeasetelma tuottaa toisen tyypin sensuroitua dataa, saadaan λ :n estimaatiksi

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^r t_i + (n-r)t_r}.$$

3.5 Lambdan ja MTTF:n luottamusvälit

Toisen tyyppin sensuroidulle datalle voidaan määrittää λ :n $100 \times (1 - \alpha)$ %:n luottamusväli seuraavasti: Olkoon r vikaantumisten lukumäärä. Tällöin χ^2 -jakauman, $\text{Khi2}(2r)$, $100 \times \alpha/2$ -prosenttipiste jaettuna vapausasteilla $2r$, tuottaa kertoimen $k_{r;\alpha/2}$. Kertomalla estimaatti, $\hat{\lambda}$, $k_{r;\alpha/2}$:lla, saadaan luottamusvälin alaraja $\lambda_{100 \times \alpha/2}$. Saman χ^2 -jakauman ylempi $100 \times (1 - \alpha/2)$ -prosenttipiste jaettuna $2r$:llä tuottaa kertoimen $k_{r;1-\alpha/2}$. Luottamusvälin yläraja saadaan lausekkeesta

$$k_{r;1-\alpha/2} \times \hat{\lambda} = \lambda_{100 \times (1-\alpha/2)}.$$

Ensimmäisen tyyppin sensuroidulle datalle voidaan laskea tarkat luottamusvälit vain, jos vikaantuneet yksiköt on korvattu välittömästi vikaantumisen jälkeen uusilla. Siinä tapauksessa luottamusvälin alaraja on täsmälleen sama kuin toisen tyyppin sensuroidulle datalle. Luottamusvälin ylärajan kerrointa k laskettaessa käytetään sen sijaan χ^2 -jakauman, $\text{Khi2}(2 \times (r + 1))$, $100 \times (1 - \alpha/2)$ -prosenttipistettä jaettuna $2r$:llä. Toisin sanoen vain χ^2 -jakauman vapausasteet ovat muuttuneet. Kertoimet $k_{r;\alpha/2}$ ja $k_{r;1-\alpha/2}$ riippuvat vain vikaantuneiden yksiköiden lukumäärästä r .

Jos yksikään testiyksikkö ei ole vikaantunut kokeen aikana, on λ :n pisteestimaatti nolla. Nolla ei ole realistinen estimaatti λ :lle, sillä se ei ota huomioon edes yksiköiden lukumäärää n . Yläraja λ :n $100 \times (1 - \alpha)$ %:n luottamusvälille on

$$\lambda_{100 \times (1-\alpha/2)} = \frac{\chi_{2;100 \times (1-\alpha/2)}^2}{2nT},$$

missä $\chi_{2;100 \times (1-\alpha/2)}^2$ on χ^2 -jakauman $100 \times (1 - \alpha)$ -prosenttipiste kahdella vapausasteella.

MTTF:n luottamusvälit saadaan muuntamalla λ :n vastaavat rajat käännteislukuun. Tällöin on huomattava, että λ :n luottamusvälin ylärajan käännteisluku on MTTF:n luottamusvälin alaraja. Vastaavasti, λ :n luottamusvälin alarajan käännteisluku on MTTF:n luottamusvälin yläraja.

3.6 Koesuunnittelu eksponenttijakauman tapauksessa

Eksponenttijakauman tapauksessa otoskoko saadaan määritettyä, kun tunnetaan vikaantumistaajuus, jota tavoitellaan, luotettavuustaso, jolla vikaantumistaajuustavoite saavutetaan, kokeen kokonaiskesto ja vikaantumisten lukumäärä, joka kokeen aikana voidaan sallia. Tällöin yhtälöstä

$$(3.2) \quad \frac{r}{n \times T} \times k_{r;1-\alpha} = \lambda_{\text{tavoite}}$$

tunnetaan n :ää lukuunottamatta kaikki suureet. Kun yhtälö ratkaistaan n :n suhteen, saadaan vaadittava otoskoko. Vastaavasti yhtälöstä voidaan ratkaista kokeen kesto T .

Koetta suunniteltaessa kannattaa päättää etukäteen millä ehdoilla koe voidaan hyväksyä ja koska se tulisi hylätä. On siis päätettävä kuinka moni koeyksikkö saa vikaantua kokeen aikana, ennen kuin koko koe hylätään. Jos otoskoko, kokeen kesto ja vikaantumistaajuustavoite luottamustasoineen on asetettu, voidaan $r \times k_{r;1-\alpha}$ ratkaista yhtälöstä (3.2). Kerrointa k ei voida ratkaista ennen kuin r tunnetaan, joten tulon tekijät on ratkaistava yhdessä. Tämä voidaan tehdä sopivalla algoritmilla, kokeilemalla tai valmiiden taulukoiden avulla (Tobias & Trindade 1986).

Minimiotoskoon määrittämiseksi täytyy tuntea vikaantumistaajuustavoite luottamustasoineen sekä kokeen kesto. Koska kyseessä on otoskoon minimi, hyväksytään koe vain, jos yksikään koeyksikkö ei vikaannu eli $r = 0$. Otokoko saadaan yhtälöstä

$$n = \frac{-\ln \alpha}{\lambda_{\text{tavoite}} \times T}.$$

Kuten otoskoon minimiä määritettäessä myös kokeen keston minimiä määritettäessä vikaantumistaajuustavoite halutulla luottamustasolla jää täyttymättä, jos yksikin koeyksikkö vikaantuu kokeen aikana. Kokeen keston minimi voidaan laskea, kun tunnetaan otoskoko sekä vikaantumistaajuustavoite luottamustasoineen. Kokeen keston minimiksi saadaan

$$T = \frac{-\ln \alpha}{n \times \lambda_{\text{tavoite}}}.$$

Minimiotoskoon käyttö on perusteltua tilanteissa, joissa ollaan vakuuttuneita, että tuote on joko selvästi parempi tai selvästi huonompi kuin kokeen vikaantumistaajuustavoite edellyttäisi. Tällöin pyritään vain vahvistamaan ennakokäsitys. Suuremman epävarmuuden vallitessa kannattaa otoskokoa kasvattaa ja sallia joitakin vikaantumisia ennen kuin koe hylätään.

3.7 Simuloitu esimerkkitapaus

3.7.1 Satunnaisotos eksponenttijakaumasta

Tarkastellaan esimerkkiaineistoa, joka on tuotettu generoimalla sadan havainnon satunnaisotos eksponenttijakaumasta. Esimerkkiaineisto vastaa koetilannetta, jossa sataa laitetta on käytetty kunnes jokainen niistä on vikaantunut. Olkoon T vikaantumisaika eli aika, joka kuluu laitteen käynnistämisestä sen vikaantumiseen. Nyt $T \sim \text{Exp}(\lambda)$. Kaikkien laitteiden tarkat vikaantumisaajat oletetaan tunnetuiksi ja ajat on annettu taulukossa 3.1. Aikayksikkö on määrittelämätön ja taulukon arvot on katkaistu kahden desimaalin tarkkuuteen. Kaikki laskenta on suoritettu R-ohjelman versiolla 1.9.1 ja R-koodi kommentteineen löytyy liiteosiosta.

3.7.2 χ^2 -yhteensopivuustesti

Tilastollisen mallin valinnassa oletetaan usein, että otos on poimittu populaatiosta, jonka jakauma tunnetaan. On hyödyllistä tarkistaa, että otos myös

Taulukko 3.1. Sadan vikaantumisajan satunnaisotos eksponenttijakaumasta.

0.02	0.13	0.29	0.44	0.45	0.46	0.47	0.51	0.52	0.55
0.79	0.81	0.92	0.93	1.09	1.14	1.29	1.43	1.47	1.61
1.63	1.85	1.93	2.00	2.18	2.20	2.21	2.43	2.46	2.52
2.62	2.64	2.71	2.92	3.03	3.03	3.43	3.59	3.59	3.80
3.85	4.11	4.26	4.43	4.59	5.00	5.35	5.37	5.43	5.90
6.02	6.05	6.74	7.18	7.25	7.29	7.36	7.44	7.52	8.01
8.43	8.74	8.87	9.05	9.45	10.25	10.38	10.92	10.97	11.14
11.18	11.42	11.46	12.36	12.57	12.58	13.20	13.55	14.58	14.80
15.29	15.73	15.85	18.32	18.71	19.00	19.75	20.36	22.14	23.53
23.75	25.22	25.35	26.75	26.84	27.35	27.41	27.41	38.54	41.92

todellisuudessa sopii yhteen oletetun jakauman kanssa. Selvissä tapauksissa visuaalinen tarkastelu riittää, mutta usein turvaudutaan esimerkiksi χ^2 -yhteensovituustestiin.

Nollahypoteesina on, että malli on sopiva. Luottamustason valinnalla määritetään, kuinka voimakkaasti otoksen on poikettava oletetusta jakaumasta ennen kuin nollahypoteesi hylätään. Jos luottamustasoksi valitaan 95 %, hylätään nollahypoteesi vain siinä tapauksessa, että näin poikkeavaan otokseen päädytään harvemmin kuin 5 %:ssa tapauksista, kun malli on oikea. Käyttökelpoisen mallin hylkääminen tunnetaan ykköstyypin virheenä (Type I error). Luottamustason valinnalla säädellään siis todennäköisyyttä tehdä ykköstyypin virhe, todennäköisyys on pienempi kuin $0.01 \times (100 - \text{luottamustaso})$.

χ^2 -yhteensovituustestiä voidaan soveltaa mille hyvänsä jakaumalle $F_T(t; \boldsymbol{\theta})$. Testin aluksi data ryhmitellään sopiviin osaväleihin. Osavälien ei tarvitse olla saman mittaisia. Jos mallissa on tuntemattomia parametreja, ne estimoidaan suurimman uskottavuuden menetelmällä. Eksponenttijakauman tapauksessa ainoa tuntematon parametri on λ , jolle saadaan suurimman uskottavuuden estimaatti yhtälön (3.1) avulla.

Kun malli on täysin määritelty, lasketaan mallin mukaisen kertymäfunktion avulla vikaantumisen todennäköisyys kullakin osavälillä. Osavälille (a_{i-1}, a_i) todennäköisyys saadaan yksinkertaisesti erotuksesta $p_i = F_T(a_i; \boldsymbol{\theta}) - F_T(a_{i-1}; \boldsymbol{\theta})$. Vikaantumisten odotettu frekvenssi kullakin osavälillä, e_i , saadaan kertomalla edellä lasketut todennäköisyydet otoskoolla

$$e_i = n \times p_i.$$

Osavälejä saatetaan joutua yhdistämään muiden osavälien kanssa, sillä vikaantumisten odotetun frekvenssin on oltava vähintään viisi kullakin osavälillä. Jos vikaantumisten havaittu frekvenssi i :nnellä osavälillä on o_i ja osavälit on valittu siten että $p_i > 0, \forall p_i$, saadaan χ^2 -testisuureen arvo lausekkeesta

$$W = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i},$$

missä k on osavälien lukumäärä. Testisuure lähestyy χ^2 -jakaumaa vapausastein $k - 1$, kun $n \rightarrow \infty$.

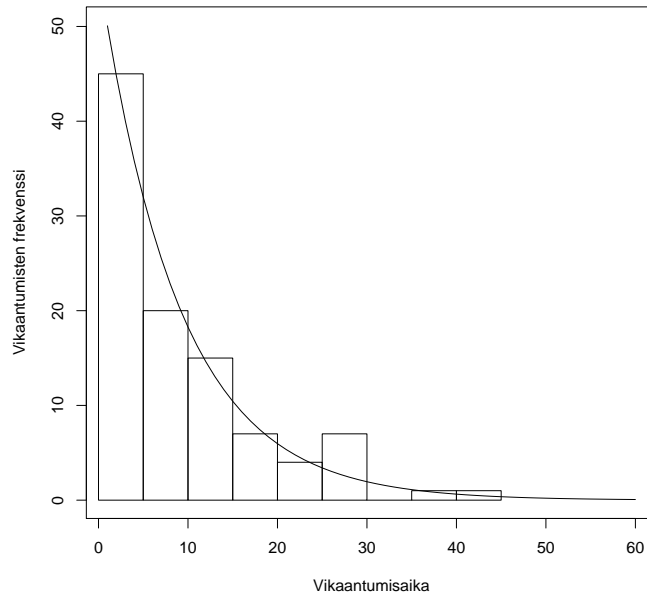
Jos parametrien arvoja joudutaan estimoimaan datasta, on W asymptoottisesti χ^2 -jakautunut kuten edellä, mutta vapausasteet vähenevät yhdellä kutakin estimoitua parametria kohti. Oletetaan, että $\boldsymbol{\theta} = \theta_1, \dots, \theta_s$ ja $\hat{p}_i = F_T(a_i; \hat{\boldsymbol{\theta}}) - F_T(a_{i-1}; \hat{\boldsymbol{\theta}})$, missä $\hat{\boldsymbol{\theta}}$ on $\boldsymbol{\theta}$:n suurimman uskottavuuden estimaatti. Tällöin

$$\hat{W} = \sum_{i=1}^k \frac{(o_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

on asymptoottisesti χ^2 -jakautunut vapausastein $k - s - 1$.

3.7.3 Esimerkki

Esimerkkiaineiston tapauksessa λ :n suurimman uskottavuuden estimaatti on $\hat{\lambda} \approx 0.112$. Kuvassa 3.1 vikaantumisaikojen histogrammia verrataan eksponenttijakauman tiheysfunktion kuvaajaan. Kuvan tiheysfunktio on muodostettu sijoittamalla λ :n estimaatti 0.112 eksponenttijakauman tiheysfunktion kaavaan. Tiheysfunktion $f(t)$ arvot on vielä kerrottu 500:lla, koska tiheysfunktion kuvaajan ja x-akselin väliin jäävä ala on yksi, ja histogrammin ala on $n \times$ pylvään leveys eli tässä tapauksessa $100 \times 5 = 500$. Näin kuvaajat on saatu samaan mittakaavaan, jolloin niiden muodon suora vertaaminen on mahdollista. Tiheysfunktion kuvaaja seuraa hyvin histogrammin muotoja, joten oletus eksponenttijakauman käyttökelpoisuudesta populaation mallina on järkevä.



Kuvio 3.1. Verrataan eksponenttijakauman tiheysfunktion kuvaajaa esimerkkiaineiston histogrammiin.

Esimerkkiaineiston osavälijako löytyy taulukosta 3.2. Taulukossa 3.2 on annettu myös vikaantumisten havaitut frekvenssit, vikaantumisen todennäköisyys sekä vikaantumisten odotetut frekvenssit kullakin osavälillä. χ^2 -yhteensopivuustestin testisuure esimerkkiaineistolle on $\hat{W} \approx 3.161$. Testisuureen arvo jää alle χ^2 -jakauman 65. prosenttipisteen, joten mallia ei ole syytä hylätä. Tulos on odotettu, sillä aineisto oli satunnaisotos $Exp(0.1)$ -jakaumasta.

Taulukko 3.2. χ^2 -yhteensopivuustestin osavälit arvoineen.

osaväli (a_{i-1}, a_i)	havaitut vikaantumiset	todennäköisyydet $F(a_i) - F(a_{i-1})$	odotetut vikaantumiset
(0,5)	38	0.393	39.3
(5,10)	28	0.239	23.9
(10,15)	10	0.145	14.5
(15,25)	13	0.141	14.1
>25	11	0.082	8.24

4 Normaalijakauma ja lognormaalijakauma

4.1 Normaalijakauman perusteita

Normaalijakauman tiheysfunktion kuvaaja, kellokäyrä, on jokaiselle tilastotiedettä opiskelleelle tuttu. Normaalijakauma ei ole aito luotettavuusjakauma, sillä käyrä on määritelty koko x-akselilla välillä $]-\infty, +\infty[$. Yleensä luotettavuusjakaumat on määritelty vain positiivisella x-akselilla. Normaalijakaumaa kuitenkin käytetään vikaantumisprosessien mallina, koska usein todennäköisyys, että normaalijakautunut satunnaismuuttuja saa negatiivisia arvoja, on hyvin pieni (Ebeling 1997). Normaalijakaumaa käytetään yleisesti esimerkiksi mittausvirheiden mallintamiseen. Normaalijakauman tiheysfunktion yhtälö on

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

missä μ on jakauman keskiarvo ja σ^2 on jakauman varianssi, varianssin neliöjuuri, σ , on hajonta. Parametrit μ ja σ^2 voidaan määrittää momenttien avulla (Liski 2005). Satunnaismuuttujan X ensimmäinen momentti eli odotusarvo on

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Satunnaismuuttujan X toinen keskusmomentti eli varianssi on

$$\begin{aligned} \sigma^2 &= E[(X - \mu)^2] \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx. \end{aligned}$$

Normaalijakauma on symmetrinen μ :n suhteen ja σ kertoo kuinka lähelle keskiarvoa alue kuvaajan alapuolella sijoittuu. Normaalijakautuneen populaation havainnoista noin 68 % sijaitsee korkeintaan $\pm\sigma$:n etäisyydellä keskiarvosta, 95 % sijaitsee korkeintaan $\pm 2\sigma$:n etäisyydellä keskiarvosta ja 99.7 % sijaitsee korkeintaan $\pm 3\sigma$:n etäisyydellä keskiarvosta.

Jos satunnaismuuttuja $X \sim N(\mu, \sigma^2)$, niin muunnoksella

$$Z = \frac{X - \mu}{\sigma}$$

saadaan satunnaismuuttuja Z , joka noudattaa standardimuotoista normaalijakaumaa, $Z \sim N(0, 1)$. Standardimuotoisen normaalijakauman tiheysfunktio on

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

ja kertymäfunktio

$$F(z) = \int_{-\infty}^z f(z) dz.$$

Z :n tiheysfunktioista käytetään usein merkintää $\phi(z)$ ja kertymäfunktioista merkintää $\Phi(z)$.

Olkoon X ja Y riippumattomia satunnaismuuttujia ja $X \sim N(\mu_x, \sigma_x^2)$ ja $Y \sim N(\mu_y, \sigma_y^2)$. Tällöin summa $W = X + Y$ noudattaa myös normaalijakaumaa odotusarvonaan $\mu_w = \mu_x + \mu_y$ ja varianssinaan $\sigma_w^2 = \sigma_x^2 + \sigma_y^2$. Tämä ominaisuus on yleistettävissä myös useammalle riippumattomalle normaalijakautuneelle satunnaismuuttujalle.

Riippumattomien normaalijakautuneiden satunnaismuuttujien lineaarikombinaatiot noudattavat myös normaalijakaumaa. Olkoot X ja Y riippumattomia satunnaismuuttujia ja $X \sim N(\mu_x, \sigma_x^2)$ ja $Y \sim N(\mu_y, \sigma_y^2)$. Tällöin $Z = aX + bY$ noudattaa myös normaalijakaumaa odotusarvonaan $(a\mu_x + b\mu_y)$ ja varianssinaan $(a^2\sigma_x^2 + b^2\sigma_y^2)$.

Olkoon x_1, x_2, \dots, x_n n :n alkion otos normaalijakaumasta $N(\mu, \sigma^2)$. Tällöin otoskeskiarvo $\bar{x} = \sum_{i=1}^n x_i/n$ noudattaa myös normaalijakaumaa odotusarvonaan $\mu_{\bar{x}} = \mu$ ja varianssinaan $\sigma_{\bar{x}}^2 = \sigma^2/n$.

4.2 Keskeinen rajaväittäjä

Keskeisen rajaväittäjän mukaan suuren riippumattomien satunnaismuuttujien joukon summa noudattaa likimain normaalijakaumaa, jos kunkin satunnaismuuttujan vaikutus yhteenlaskun lopputulokseen on vähäinen. Satunnaismuuttujien määrän lisääminen vähentää yksittäisen muuttujan vaikutusta kokonaisuuteen, jolloin summa noudattaa yhä tarkemmin normaalijakaumaa. Täsmällisempi keskeisen rajaväittäjän määritelmä löytyy esimerkiksi Liskin matemaattisen tilastotieteen opetusmonisteesta (2005, s. 223–225).

Otoskeskiarvo on hyvä esimerkki tilanteesta, johon keskeistä rajaväittäjää voidaan soveltaa. Olkoon x_1, x_2, \dots, x_n otos jakaumasta, jonka keskiarvo ja varianssi ovat äärelliset. Summan $\bar{x} = (x_1/n) + (x_2/n) + \dots + (x_n/n)$ jokainen tekijä pienenee n :n kasvaessa, ja \bar{x} :n jakauma lähenee normaalijakaumaa, riippumatta siitä, mitä jakaumaa x_i :t noudattavat. Otoskeskiarvo \bar{x} noudattaa yleensä melko tarkasti normaalijakaumaa jo niinkin pienillä n :n arvoilla kuin neljä tai viisi (Tobias & Trindade 1986).

Normaalijakauman laaja soveltuvuus satunnaisilmiöiden malliksi perustuu juuri keskeiseen rajaväittäjään. Satunnaisilmiöitä voidaan käsitellä pienten osatekijöiden summina. Esimerkiksi mittausvirheet syntyvät monien pienten

satunnaisten tekijöiden vaikutuksesta, ja yhdessä nämä tekijät muodostavat poikkeaman todellisesta mittausarvosta.

4.3 Normaalijakauman parametrien estimointi

Sensuroimattoman datan tapauksessa normaalijakauman parametrien μ ja σ^2 suurimman uskottavuuden estimaattorit saadaan helposti ratkaistua. Olkoon y_1, y_2, \dots, y_n satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$, jossa parametrien otosavaruus on

$$\Omega = (\mu, \sigma^2) : -\infty < \mu < \infty, 0 < \sigma^2 < \infty.$$

Tällöin uskottavuusfunktio on

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y_i - \mu)^2}{2\sigma^2}\right].$$

Uskottavuusfunktio voidaan kirjoittaa yhtäpitävästi muodossa

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left[\frac{-\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right], \quad (\mu, \sigma^2) \in \Omega.$$

Logaritmoitu uskottavuusfunktio on

$$l(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}.$$

Osittaisderivaatat μ :n ja σ^2 :n suhteen ovat

$$\frac{\partial(\ln L)}{\partial\mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu)$$

ja

$$\frac{\partial(\ln L)}{\partial\sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2.$$

Yhtälöllä $\partial(\ln L)/\partial\mu = 0$ on ratkaisu $\mu = \bar{y}$. Kun yhtälöä $\partial(\ln L)/\partial\sigma^2 = 0$ ratkaistaessa μ korvataan \bar{y} :llä, saadaan ratkaisuksi

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Toisten osittaisderivaattojen perusteella voidaan päätellä, että ratkaisut ovat maksimeja. Täten μ :n ja σ^2 :n suurimman uskottavuuden estimaattorit ovat

$$\hat{\mu} = \bar{y} \quad \text{ja} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Vaihtoehtoinen estimaattori σ^2 :lle on otosvarianssi $s^2 = \sum (y_i - \bar{y})^2 / (n - 1)$, joka on σ^2 :n harhaton estimaattori. Estimaattorit \bar{y} ja s^2 voidaan osoittaa μ :n ja

σ^2 :n minimivarianssisiksi harhattomiksi estimaattoreiksi (Lawless 1982). (Hogg & Tanis 1988.)

Suurimman uskottavuuden estimaattoreiden löytäminen sensuroidulle datalle ei ole yhtä yksinkertaista. Merkitään standardimuotoisen normaalijakauman tiheysfunktiota $\phi(z)$:llä ja luotettavuusfunktiota $R(z)$:lla,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{ja} \quad R(z) = \int_z^\infty \phi(x) dx, \quad -\infty < z < \infty.$$

Jos $Y \sim N(\mu, \sigma^2)$, niin Y :n tiheysfunktio on $\sigma^{-1}\phi[(y - \mu)/\sigma]$ ja luotettavuusfunktio $R[(y - \mu)/\sigma]$. Oletetaan, että otoksessa on n :n laitteen vikaantumisaajat, ja otos voi olla joko ensimmäisen tai toisen tyyppin sensuroidusta datasta. Merkitään sekä vikaantumisaikoja että sensurointiaikoja y_i :llä, ($i = 1, \dots, n$). Olkoon D niiden aikojen joukko, joille y_i on havaittu vikaantumisaika, ja olkoon C niiden aikojen joukko, joille y_i on sensurointiaika. Havaittujen vikaantumisaikojen lukumäärä on $r = |D|$. Uskottavuusfunktio on

$$L(\mu, \sigma^2) = \prod_{i \in D} \frac{1}{\sigma} \phi\left(\frac{y_i - \mu}{\sigma}\right) \prod_{i \in C} R\left(\frac{y_i - \mu}{\sigma}\right).$$

Logaritmoitu uskottavuusfunktio on

$$\ln L(\mu, \sigma^2) = -r \ln \sigma - \frac{1}{2\sigma^2} \sum_{i \in D} (y_i - \mu)^2 + \sum_{i \in C} \ln R\left(\frac{y_i - \mu}{\sigma}\right).$$

Logaritmoidun uskottavuusfunktion ensimmäiset osittaisderivaatat ovat

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i \in D} (y_i - \mu) + \frac{1}{\sigma} \sum_{i \in C} \phi\left(\frac{y_i - \mu}{\sigma}\right) / R\left(\frac{y_i - \mu}{\sigma}\right) \\ \frac{\partial \ln L}{\partial \sigma} &= -\frac{r}{\sigma} + \frac{1}{\sigma^3} \sum_{i \in D} (y_i - \mu)^2 + \frac{1}{\sigma} \sum_{i \in C} \frac{y_i - \mu}{\sigma} \phi\left(\frac{y_i - \mu}{\sigma}\right) / R\left(\frac{y_i - \mu}{\sigma}\right). \end{aligned}$$

Suurimman uskottavuuden yhtälöt saadaan asettamalla $\partial \ln L / \partial \mu = 0$ ja $\partial \ln L / \partial \sigma = 0$. Näiden yhtälöiden käsittelyä voidaan helpottaa määrittelemällä $z_i = (y_i - \mu)/\sigma$, ($i = 1, 2, \dots, n$), ja

$$V(z) = -\frac{d}{dz} \ln R(z) = \frac{\phi(z)}{R(z)}$$

$$\lambda(z) = \frac{d}{dz} V(z) = V(z)[V(z) - z],$$

missä $V(z)$ on normaalijakauman hasardifunktio ja $\lambda(z)$ hasardifunktion derivaatta. $V(z)$:n ja $\lambda(z)$:n avulla voidaan μ :n ja σ :n suurimman uskottavuuden yhtälöt esittää muodossa

$$(4.1) \quad \sum_{i \in D} z_i + \sum_{i \in C} V(z_i) = 0$$

$$(4.2) \quad -r + \sum_{i \in D} z_i^2 + \sum_{i \in C} z_i V(z_i) = 0.$$

Logaritmoidun uskottavuusfunktion $\ln L$ toiset derivaatat ovat

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial \mu^2} &= -\frac{r}{\sigma^2} - \frac{1}{\sigma^2} \sum_{i \in C} \lambda(z_i) \\ \frac{\partial^2 \ln L}{\partial \sigma^2} &= \frac{r}{\sigma^2} - \frac{3}{\sigma^2} \sum_{i \in D} z_i^2 - \frac{2}{\sigma^2} \sum_{i \in C} z_i V(z_i) - \frac{1}{\sigma^2} \sum_{i \in C} z_i^2 \lambda(z_i) \\ \frac{\partial^2 \ln L}{\partial \mu \partial \sigma} &= -\frac{2}{\sigma^2} \sum_{i \in D} z_i - \frac{1}{\sigma^2} \sum_{i \in C} V(z_i) - \frac{1}{\sigma^2} \sum_{i \in C} z_i \lambda(z_i).\end{aligned}$$

Suurimman uskottavuuden yhtälöistä (4.1) ja (4.2) voidaan ratkaista $\hat{\mu}$ ja $\hat{\sigma}$ Newton-Raphson-menetelmällä. Toisinaan algoritmi ei konvergoi ilman hyviä lähtöarvoja. Lähtöarvoina voidaan käyttää esimerkiksi μ :n ja σ :n parhaita lineaarisia harhattomia estimaatteja (BLUE). (Lawless 1982.)

4.4 Luottamusväli μ :lle ja σ :lle

Sensuroimattoman datan tapauksessa μ :n luottamusväli voidaan estimoida napasuureen $Z_1 = (\bar{y} - \mu)\sqrt{n}/s$ avulla, missä $\bar{y} = \sum y_i/n$ ja $s^2 = \sum (y_i - \bar{y})^2/(n-1)$ (Lawless 1982). Jos y_1, y_2, \dots, y_n on otos normaalijakaumasta, niin Z_1 noudattaa t-jakaumaa vapausastein $n-1$ ($Z_1 \sim t_{n-1}$). Valitaan $t_{\alpha/2; n-1}$ siten että $P[Z_1 \geq t_{\alpha/2; n-1}] = \alpha/2$. Silloin

$$\begin{aligned}1 - \alpha &= P\left[-t_{\alpha/2; n-1} \leq \frac{\bar{y} - \mu}{s/\sqrt{n}} \leq t_{\alpha/2; n-1}\right] \\ &= P\left[-t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right) \leq \bar{y} - \mu \leq t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right)\right] \\ &= P\left[-\bar{y} - t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right) \leq -\mu \leq -\bar{y} + t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right)\right] \\ &= P\left[\bar{y} - t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \bar{y} + t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right)\right].\end{aligned}$$

Täten

$$\left[\bar{y} - t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right), \bar{y} + t_{\alpha/2; n-1} \left(\frac{s}{\sqrt{n}}\right)\right]$$

on μ :n $100(1 - \alpha)$ %:n luottamusväli. (Hogg & Tanis 1988.)

Vastaavasti saadaan suureen $Z_2 = (n-1)s^2/\sigma^2$ avulla muodostettua luottamusväli σ :lle (Lawless 1982). Jos y_1, y_2, \dots, y_n on satunnaisotos normaalijakaumasta, niin Z_2 noudattaa χ^2 -jakaumaa vapausastein $n-1$ ($Z_2 \sim \chi_{n-1}^2$).

4.5 Lognormaalijakauma

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

$$t_f = e^X$$

noudattaa lognormaalia jakaumaa parametrein $T_{50} = e^\mu$ ja σ . Kääntäen, lognormaalijakautuneiden vikaantumisaikojen logaritmit noudattavat normaalijakaumaa. Jos alkutilanteena on satunnaisten vikaantumisaikojen, t_f , populaatio, jota mallinnetaan lognormaalijakaumalla parametrein T_{50} ja σ , on vikaantumisaikojen luonnollisten logaritmien populaatio normaalijakautunut keskiarvonaan $\mu = \ln T_{50}$ ja keskihajontanaan σ . Muunnoksen helppouden ansiosta lognormaalijakautunutta dataa ei tarvitse käsitellä sellaisenaan, vaan muunnettua dataa voidaan analysoida normaalijakauman menetelmin.

T_{50} on lognormaalijakauman mediaani, σ ei kuitenkaan ole lognormaalijakauman keskihajonta, vaan ainoastaan muotoparametri. Lognormaalijakauman parametrina σ :a käytetään, koska sen käyttö on matemaattisesti edullista (Tobias & Trindade 1986).

Lognormaalijakauman tiheysfunktio on

$$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} e^{-(1/2\sigma^2)(\ln t - \ln T_{50})^2}.$$

Lognormaalijakauman kertymäfunktio voidaan esittää standardimuotoisen normaalijakauman kertymäfunktion avulla

$$F(t) = \int_0^t f(u) du = \Phi \left[\frac{\ln(t/T_{50})}{\sigma} \right].$$

Odotusarvo on

$$E(t_f) = MTTF = T_{50} e^{\sigma^2/2}$$

ja varianssi

$$Var(t_f) = T_{50}^2 e^{\sigma^2} (e^{\sigma^2} - 1).$$

Lognormaalijakauman tiheysfunktion kuvaajan muoto vaihtelee voimakkaasti muotoparametrin σ mukana. Muotojen kirjo muistuttaa Weibull-jakauman tiheysfunktion kuvaajan mahdollisia muotoja. Lognormaalijakauma ja Weibull-jakauma soveltuvatkin usein yhtä hyvin saman havaintojoukon malleiksi. Ei ole kuitenkaan yhdentekevää kumpaa jakaumaa käyttää, sillä niiden väliltä löytyy myös eroavaisuuksia. Esimerkiksi jakaumien teoreettinen tausta on erilainen: Weibull-jakauma voidaan johtaa ääriarvojakaumana, joka toimii, kun useat pienet viat kilpailevat keskenään siitä, mikä niistä ensimmäisenä ehtii aiheuttaa koko laitteen vikaantumisen. Lognormaalijakauma puolestaan toimii mallina prosesseille, joissa ajan myötä heikkenevän laadun — esimerkiksi materiaalin väsymisen — takia ajaudutaan kohti vikaantumista.

4.6 Lognormaalijakauman parametrien estimointi

Jos käytössä on tarkkojen vikaantumisaikojen sensuroimaton otos, voidaan T_{50} ja σ estimoida ottamalla ensin kaikista vikaantumisajoista luonnollinen loga-

ritmi ja laskemalla sitten otoskeskiarvo \bar{x} ja otoshajonta s logaritmoidusta datasta. Tarvittavat lausekkeet ovat

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n \ln t_i}{n} \\ s &= \sqrt{\frac{\sum_{i=1}^n (\ln t_i - \bar{x})^2}{n-1}} \\ \hat{T}_{50} &= e^{\bar{x}} \\ \hat{\sigma} &= s.\end{aligned}$$

Sensuroidulle tai ryhmitellylle datalle nämä estimaattorit eivät sovellu. Tällöin suurimman uskottavuuden estimaatit on ratkaistava numeerisin menetelmin.

4.7 Esimerkki lognormaalijakautuneella datalla

4.7.1 Paukkumaissaineisto

Aineiston hankkimiseksi paukkumaissia kuumennettiin uunissa, uunipellillä, kolmessa eri lämpötilassa: 200, 225 ja 250 Celsius-asteessa. Kunkin maissinjyvän osalta mitattiin aika uuniin laitosta jyvän poksauttamiseen. Jyvät laitettiin uuniin kymmenen kappaleen erissä, jotta ajat pystyttiin havainnoimaan riittävän tarkasti tavallisissa keittiöolosuhteissa. Uunin oikean lämpötilan varmistamiseksi käytettiin uunilämpömittaria. Niiden jyvien osalta, jotka eivät kohutuullisessa ajassa poksautaneet auki, on käytössä vain sensurointiaika. Tällaisia aukeamattomia jyviä oli vain kolme ja ne kaikki olivat 200-asteisessa uunissa. Kustakin lämpötilasta mitattiin sata poksahdusaikaa ja 200-asteisessa uunissa lisäksi kolme sensurointiaikaa. Kyseessä on toisen tyyppin sensuroitu data, ja poksahdusaikoja voidaan käsitellä kuten vikaantumisaikoja. Esimerkin R-koodi löytyy liiteosiosta.

Kaikkien kolmen lämpötilan osalta piirrettiin histogrammi jakauman muodon määrittämiseksi. Histogrammit muistuttivat tunnetuista tavallisista jakaumista muodoltaan eniten Weibull- ja lognormaalijakaumia. Lognormaalijakaumaan päädyttiin, koska paukkujyvien poksahdusmekanismi muistuttaa enemmän lognormaalijakauman kuin Weibull-jakauman teoreettista perustaa. Lognormaalijakauma toimii mallina esimerkiksi silloin, kun materiaalin väsyminen aiheuttaa vikaantumisen. Maissinjyvä poksauttaa, kun jyvän sisäinen paine kasvaa niin suureksi, että kuori räjähtää auki. Jyvä sisältää hieman vettä, joka kuumetessaan laajenee ja aiheuttaa paineen.

4.7.2 Sensuroimattoman datan parametrien estimaatit

Käsitellään tässä vaiheessa aineistoa kuin se olisi sensuroimatonta, eli jätetään kolme sensurointiaikaa kokonaan tarkastelun ulkopuolelle. Tällöin käytössä on kolme sadan vikaantumisajan otosta. Muunnetaan ensin otokset normaalijakautuneiksi ottamalla kaikista vikaantumisajoista luonnollinen logaritmi. Estimoidaan normaalijakautuneiden otosten parametrit kuten kappaleessa 4.3.

Estimoinnin tulokset ovat taulukossa 4.1. Estimaateista nähdään, että lämpötilan kasvaessa sekä otoskeskiarvo että otosvarianssi pienenevät. Varianssien luottamusvälit menevät tosin osittain päällekkäin.

Otosjakaumien normaalisuusoletus testattiin Shapiro-Wilk-testillä. Testisuureiden arvot 200-, 225- ja 250-asteessa olivat vastaavasti $W = 0.9918$, p -arvo = 0.8064, $W = 0.9848$, p -arvo = 0.3067 ja $W = 0.9878$, p -arvo = 0.4925. Normaalisuusoletus jää kaikkien otosten osalta voimaan.

Taulukko 4.1. Parametrien estimaatit luottamusväleiseen normaalijakautuneille otoksille.

lämpötila	$\hat{\mu}$	$\hat{\mu}$:n 95%:n luottamusväli	$\hat{\sigma}^2$	$\hat{\sigma}^2$:n 95%:n luottamusväli
200	5.041	(5.000,5.084)	0.045	(0.035,0.061)
225	4.493	(4.455,4.531)	0.036	(0.028,0.049)
250	4.372	(4.339,4.405)	0.027	(0.021,0.037)

Lognormaalijakauman mediaanin estimaatti saadaan normaalijakauman keskiarvosta muunnoksella $\hat{T}_{50} = e^{\bar{x}}$. Esimerkkiaineiston tapauksessa lognormaalijakautuneiden otosten mediaanit ovat $\hat{T}_{50}(200) = 154.6$, $\hat{T}_{50}(225) = 89.4$ ja $\hat{T}_{50}(250) = 79.2$, missä sulkeissa oleva luku on uunin lämpötila ja mittayksikkö on sekunti. 200-asteisessa uunissa mediaaniaika on siis noin 2 minuuttia 30 sekuntia ja 250-asteisessä uunissa noin puolet tästä.

5 Coxin regressio ja nopeutettu vikaantuminen

5.1 Verrannollisten hasardien malli

Verrannollisten hasardien malleissa eri tarkastelukohteiden hasardifunktioiden suhde ei riipu ajasta. Toisin sanoen suhde $h(t|\mathbf{x}_1)/h(t|\mathbf{x}_2)$, missä \mathbf{x}_1 ja \mathbf{x}_2 ovat regressioanalyysin selittävien muuttujien vektoreita, on ajasta riippumaton. Kun \mathbf{x} tunnetaan voidaan T :n hasardifunktio kirjoittaa muodossa

$$h(t|\mathbf{x}) = h_0(t)g(\mathbf{x}),$$

missä $h_0(t)$ on perustason hasardifunktio eli hasardifunktio tilanteessa, jossa $g(\mathbf{x}) = 1$. Sekä h_0 että g saattavat sisältää tuntemattomia parametrejä.

Luotettavuusfunktio

$$R(t|\mathbf{x}) = \exp(-H(t)) = \exp\left(-\int_0^t h(u|\mathbf{x})du\right)$$

voidaan nyt kirjoittaa muodossa

$$R(t|\mathbf{x}) = R_0(t)^{g(\mathbf{x})},$$

missä

$$R_0(t) = \exp\left(-\int_0^t h_0(u)du\right)$$

on perustason luotettavuusfunktio eli luotettavuusfunktio tilanteessa, jossa $g(\mathbf{x}) = 1$. Tästä seuraa, että verrannollisten hasardien malleissa luotettavuusfunktiot säilyttävät järjestyksensä ajan muuttuessa: $R(t|\mathbf{x}_1) \geq R(t|\mathbf{x}_2)$, tai $R(t|\mathbf{x}_1) \leq R(t|\mathbf{x}_2)$, kaikilla t :n arvoilla. (Lawless 1982.)

Suosittu ja erityisen hyödyllinen verrannollisten hasardien malli on Coxin regressiomalli, jossa $g(\mathbf{x}) = e^{\mathbf{x}'\boldsymbol{\beta}}$. Malli on siis muotoa

$$h(t|\mathbf{x}) = h_0(t)e^{\mathbf{x}'\boldsymbol{\beta}},$$

missä $\mathbf{x}'\boldsymbol{\beta} = x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p$ ja β_i :t ovat tuntemattomia regressio-kertoimia. Koska $e^{\mathbf{x}'\boldsymbol{\beta}}$ on aina positiivinen, on $h(t|\mathbf{x})$ ei-negatiivinen kaikilla \mathbf{x} :n ja $\boldsymbol{\beta}$:n arvoilla. Coxin regressiomallissa perustason hasardifunktio, $h_0(t)$, on hasardifunktio tilanteessa, jossa $\mathbf{x} = \mathbf{0}$.

5.2 Mallin parametrien estimoinnista

Coxin regressiomallissa on kaksi tuntematonta parametria: β ja $h_0(t)$. Perustason hasardifunktion sijasta mallissa voidaan käyttää perustason luotettavuusfunktia, jolloin malli voidaan esittää muodossa

$$\begin{aligned} R(t|\mathbf{x}) &= \exp\left(-\int_0^t h(u|\mathbf{x})du\right) \\ &= (R_0(t))^{\exp(\mathbf{x}'\beta)}, \end{aligned}$$

kun \mathbf{x} on annettu.

Coxin regressiomallissa vikaantumisaikojen jakaumaa ei oleteta tunnetuksi eli kyseessä on jakaumasta vapaa menetelmä. Tuntemattomat parametrit estimoidaan datasta, jossa voi olla sensuroituja vikaantumisaikoja. Analyysissä käytetään β :lle uskottavuusfunktia, joka ei riipu $h_0(t)$:sta. Näin vältetään maksimoimasta uskottavuusfunktia samanaikaisesti sekä β :n että $h_0(t)$:n tai $R_0(t)$:n suhteen. $R_0(t)$ tai $h_0(t)$ estimoidaan vasta, kun β on estimoitu.

Oletetaan, että n :n laitteen satunnaisotoksessa on k havaittua vikaantumisaikaa ja $n - k$ sensurointiaikaa. Vikaantumisaikoja merkitään $t_{(i)}$:llä ja $t_{(1)} < \dots < t_{(k)}$. Olkoon $Z_i = Z(t_{(i)})$ niiden sensuroimattomien laitteiden joukko, jotka toimivat hetkeä ennen $t_{(i)}$:tä. Uskottavuusfunktio β :lle on

$$(5.1) \quad L(\beta) = \prod_{i=1}^k \left(e^{\mathbf{x}'_i \beta} / \sum_{l \in Z_i} e^{\mathbf{x}'_l \beta} \right),$$

missä \mathbf{x}_i on selittävien muuttujien vektori laitteelle, jonka havaittu vikaantumisaika on $t_{(i)}$. Tähän uskottavuusfunktioon on päädytty tarkastelemalla todennäköisyyttä, että laite i , $i \in Z(t)$, vikaantuu hetkellä t . Olkoon $Z(t)$ annettu ja oletetaan, että hetkellä t jokin laitteista vikaantuu. Todennäköisyys, että vikaantunut laite on i , on

$$h(t|\mathbf{x}_i) / \sum_{l \in Z(t)} h(t|\mathbf{x}_l) = e^{\mathbf{x}'_i \beta} / \sum_{l \in Z(t)} e^{\mathbf{x}'_l \beta}.$$

Uskottavuusfunktio on näiden tekijöiden tulo.

Uskottavuusfunktio (5.1) ei riipu $h_0(t)$:sta ja funktio voidaan maksimoida suurimman uskottavuuden estimaattorin $\hat{\beta}$ saamiseksi. Coxin regressiomallin tapauksessa $L(\beta)$ ei ole uskottavuusfunktio uskottavuusfunktion tavanomaisessa merkityksessä ja sen käyttökelpoisuus on herättänyt keskustelua. Melko yleisten ehtojen vallitessa on kuitenkin täysin hyväksyttävää kohdella funktia (5.1) kuin tavanomaista uskottavuusfunktia (Lawless 1982).

Koska jatkuvien muuttujien arvot on usein pyöristetty sopivaan tarkkuuteen, saattaa data sisältää useita havaintoja, joilla on sama arvo. Jos samoja

arvoja on runsaasti, kannattaa analysointiin käyttää diskreettiä mallia. Jos samoja arvoja on vähän, voidaan yleensä käyttää jatkuvaa mallia, mutta tällöin uskottavuusfunktioon on tehtävä joitain muutoksia. Uskottavuusfunktio (5.1) voidaan korvata funktiolla

$$(5.2) \quad L(\boldsymbol{\beta}) = \prod_{i=1}^k e^{\mathbf{S}'_i \boldsymbol{\beta}} / \left(\sum_{l \in Z_i} e^{\mathbf{x}'_l \boldsymbol{\beta}} \right)^{d_i},$$

missä d_i on niiden vikaantumisaikojen lukumäärä, jotka ovat yhtäsuuria kuin $t_{(i)}$, ja \mathbf{S}_i on näiden d_i laitteen regressiovektoreiden, \mathbf{x} , summa. Jos siis D_i on hetkellä $t_{(i)}$ vikaantuneiden laitteiden joukko, niin $d_i = |D_i|$ ja $\mathbf{S}_i = \sum_{l \in D_i} \mathbf{x}_l$. Jos samoja arvoja ei ole, jokainen $d_i = 1$ ja (5.2) palautuu funktioksi (5.1).

5.3 Muuttujien valinnasta

Analyysin alkuvaiheessa valitaan muuttujat, joiden sopivuutta malliin tullaan koettamaan, sekä pyritään arvioimaan mahdollisten selittävien muuttujien vaikutusta vikaantumisaikoihin karkealla tasolla. Kun muuttujat on alustavasti valittu, on vielä selvitettävä käyttäytyvätkö ne mallin edellyttämällä tavalla. Jos selittävä muuttuja, x_1 , on dikotominen ja voi saada vain arvoja x_{11} tai x_{12} , voidaan data jakaa kahteen osaan muuttujan arvojen perusteella, ja sovittaa kumpaankin osadataan oma verrannollisten hasardien mallinsa. Näin saadut mallit, joiden hasardifunktiot ovat muotoa

$$h_i(t|\mathbf{x}^-) = h_{0i}(t)e^{(\mathbf{x}^-)'(\boldsymbol{\beta}^-)} \quad i = 1, 2$$

sovitetaan kumpikin omaan osadataan, joista toiselle $x_1 = x_{11}$ ja toiselle $x_1 = x_{12}$. Nyt \mathbf{x}^- on vektori (x_2, \dots, x_p) , ja $\boldsymbol{\beta}^- = (\beta_2, \dots, \beta_p)$. Coxin regressiomallin oletusten vallitessa on

$$\begin{aligned} h_i(t|\mathbf{x}^-) &= h(t|x_{1i}, \mathbf{x}^-) \\ &= h_0(t)e^{x_1 \beta_1} e^{(\mathbf{x}^-)'(\boldsymbol{\beta}^-)}. \end{aligned}$$

Jos muuttujat käyttäytyvät mallin edellyttämällä tavalla, ovat hasardifunktiot $h_{0i}(t)$ verrannollisia. Tämä voidaan varmentaa visuaalisesti piirtämällä kuvaajat

$$\ln(-\ln \hat{R}_{0i}(t)) \quad i = 1, 2,$$

missä

$$R_{0i}(t) = \exp\left(-\int_0^t h_{0i}(u) du\right)$$

on perustason luotettavuusfunktio laitteille, joilla $x_1 = x_{1i}$. Mikäli kuvaajat ovat likimain yhdensuuntaiset, ovat hasardifunktiot $h_{0i}(t)$ verrannollisia. Poikkeamat yhdensuuntaisuudesta antavat aiheita epäillä x_1 :n sopivuutta Coxin

regressiomalliin. Jos β^- estimoidaan erikseen kummallekin $i = 1, 2$, tulisi estimaattien olla likimain yhtäsuuret. Jos estimaatit eivät ole yhtäsuuret, on $x_{1:n}$ ja muiden selittävien muuttujien välillä riippuvuutta. Menetelmiä voidaan soveltaa myös muuttujille, jotka saavat useampia kuin kaksi arvoa. Tällöin data jaetaan useampaan osaan.

5.4 Kaksi esimerkkiä Coxin regressioanalyysistä

5.4.1 Alkuperäinen paukkumaissiaineisto

SPSS Advanced Models -paketti sisältää muutaman muun lisäosan ohella myös Coxin regressioanalyysin, jolla paukkumaissiaineisto on analysoitu. Analyysissä olivat mukana kaikki 300 vikaantumisaikaa sekä kolme sensurointiaikaa. Syötearvoina annettiin selitettäväksi muuttujaksi aika sekä indikaattorifunktio, joka ilmoittaa onko kyseessä vikaantumis- vai sensurointiaika. Selittävät muuttujat voivat olla sekä jatkuvia että diskreettejä. Nyt selittävänä muuttujana oli lämpötila, jota käsiteltiin kolmiarvoisena luokittelumuuttujana.

SPSS käyttää regressiokertoimien arvojen, $\hat{\beta}$, laskemiseen vain datan sensuroimattomia tapauksia. Perustason hasardifunktion, $h_0(t)$, laskennassa käytetään myös sensuroituja tapauksia. Estimoitua mallia verrataan malliin, jossa kaikki regressiokertoimet ovat nollia. Molemmista malleista lasketaan $(-2 \times \log \text{likelihood})$ ja uskottavuussuhdetestillä testataan onko jokin regressiokertoimista tilastollisesti merkitsevä. Paukkumaissiaineistolle uskottavuussuhdetestin p-arvo on pienempi kuin 0.000, joten ainakin yksi kertoimista on merkitsevä.

SPSS tulostaa regressiokertoimien arvot (B), arvojen keskivirheet sekä Waldin testisuureen arvon vapausasteineen ja p-arvoineen kullekin kertoimelle. Lisäksi tulosteessa on arvo $\exp(B)$, joka kertoo riskin muutoksen ennusteen, kun regressiokerroin kasvaa yhden yksikön verran. Jos $\exp(B)$ on lähellä ykköstä, on kertoimen vaikutus vähäinen. Haluttaessa saa tulosteeseen myös $\exp(B)$:n luottamusvälin.

Taulukko 5.1. Coxin regressioanalyysin tuloksia.

B	keskivirhe	Waldin testisuure	vapausasteet	p-arvo
		211.559	2	0.000
-2.444	0.172	201.434	1	0.000
-0.343	0.149	5.311	1	0.021

Jos selittävä muuttuja on luokitteluasteikollinen, on tulosteessa rivi luokkien yhteisvaikutukselle ja oma rivinsä kullekin luokalle, lukuunottamatta luokkaa, johon muita luokkia verrataan. Paukkumaissiaineistolle lämpötilamuuttujan luokkien yhteisvaikutuksen Waldin testisuureen arvo kahdella vapausasteella on likimain 211.6 (taulukko 5.1, 1. rivi). Tulos on erittäin merkitsevä,

joten ainakin yksi regressiokerroin on nolasta poikkeava. Vertailuluokkana on 250 astetta, joten regressiokertoimet on laskettu luokille 200 astetta (taulukko 5.1, 2. rivi) ja 225 astetta (taulukko 5.1, 3. rivi). Molempien kertoimien p-arvo on alle 0.05, joten niitä ei ole syytä poistaa mallista. Molemmat regressiokertoimet ovat negatiivisia (taulukko 5.1), mikä on luonnollista, sillä 250 asteessa paukkumaissit kypsyvät keskimäärin nopeammin kuin 200 tai 225 asteessa. Negatiiviset regressiokertoimet tarkoittavat, että vikaantumisen riski on pienempi kuin vertailuluokassa.

Taulukko 5.2. Coxin regressioanalyysin tuloksia.

luokka	$exp(B)$	$exp(B)$:n 95 %:n luottamusväli
200 astetta	0.087	(0.062,0.122)
225 astetta	0.709	(0.530,0.950)

Taulukossa 5.2 on annettu $exp(B)$ luottamusväleinen. Vertailuluokkana on edelleen 250 astetta. Myös $exp(B)$:n arvoista nähdään, että vikaantumistaajuus on 200 ja 225 asteessa pienempi kuin 250 asteessa. Jos $exp(B)$:n arvo on pienempi kuin yksi, on regressiokertoimen vaikutus hasardifunktion arvoa pienentävä. Jos $exp(B)$:n arvo on suurempi kuin yksi, on regressiokertoimen vaikutus hasardifunktion arvoa kasvattava. Jos $exp(B)$:n luottamusväli sisältää arvon 1, ei kertoimella ole vaikutusta.

SPSS tuottaa pyydettyä myös useita kuvaajia Coxin regressioanalyysiin liittyen. Aineistosta voidaan piirtää luotettavuusfunktion, kertymäfunktion ja kumulatiivisen hasardifunktion kuvaajat. Lisäksi käytettävissä on $\ln(-\ln \hat{R}(t))$:n kuvaaja eli log-miinus-log- tai lml-kuvaaja. Kuvaajat edustavat joko teoreettista keskiarvovyksilöä tai luokkakeskiarvoja. Luokittain piirrettyä lml-kuvaajaa käytetään varmentamaan verrannollisuusehdon täyttyminen. Jos $\ln(-\ln \hat{R}(t))$ -funktioiden kuvaajat ovat yhdensuuntaisia, ei ositettuun Coxin regressioon ole tarvetta. Paukkumaissiaineiston tapauksessa luokkakohtaiset lml-kuvaajat ovat likimain yhdensuuntaiset (Liite 3). Liiteosiosta löytyy lisäksi paukkumaissiaineiston kumulatiivisen hasardifunktion kuvaaja kullekin lämpötilalle.

5.4.2 Pieni paukkumaissiaineisto

Maissinjyvien hasardifunktion käytöstä arvioitiin myös pienemmän aineiston avulla. Pienemmässä aineistossa vikaantumisaikoja on vain 29 ja ne kaikki on mitattu 200-asteisessa uunissa. Yhdestä jyvistä käytössä on sensurointi-aika. Selittävinä muuttujina malliin otettiin jyvän pituus ja leveys sekä kaksiarvoinen luokittelumuuttuja, joka ilmaisee onko jyvän pinnassa vaurioita tai värimuunnoksia. Maissinjyvien pituudet otoksessa vaihtelevat välillä 6.5–10.0 mm ja leveydet välillä 5.0–8.0 mm ja ne on mitattu 0.5 mm:n tarkkuudella. Vikaantumisaikat on mitattu sekunnin tarkkuudella ja ne vaihtelevat välillä 91–165 s.

Sensuroidusta jyvän vikaantumisajasta tiedetään vain, että se kesti kuumuutta yli 180 s.

Taulukko 5.3. Coxin regressioanalyysin tuloksia pienelle aineistolle.

muuttuja	B	Waldin testisuure	vapausasteet	p-arvo
vaurio	-0.038	0.006	1	0.936
pituus	-0.278	1.603	1	0.205
leveys	-0.288	0.751	1	0.386

Analyysi toteutettiin kuten edellisessä esimerkissä. Yhdenkään regressio-kertoimen ei havaittu poikkeavan nolasta, joten valituista muuttujista ei ollut selittämään hasardifunktion kulkua. Regressiokertoimet ja Waldin testin tulokset ovat taulukossa 5.3. Jyvän koolla ja mahdollisilla pintavaurioilla oli testituloksista huolimatta vaikutusta lopputulokseen. Pienissä jyvissä on vähemmän tärkkelystä kuin suurissa jyvissä, joten auettuaan pienet jyvät ovat edelleen suurempia verrokkejaan pienempiä. Vaurioituneet jyvät eivät ehkä poksahda lainkaan vaan päästävät auetessaan vain pienen pihauksen, ja lopputulos on usein varsin mitätön niin kooltaan kuin myös ulkonäöltään.

5.5 Nopeutuskerroin

Vikaantumisdataa ei aina saada riittävästi tai riittävän nopeasti laitteiden normaaleissa käyttöoloissa. Tällöin laitteita voidaan kuormittaa tavanomaisia käyttöolosuhteita selvästi raskaammassa oloissa. Menettelyä kutsutaan nopeutetuksi elinaikatestaukseksi. Jotta nopeutetun elinaikatestauksen vikaantumisdataa voidaan hyödyntää tavallisissa käyttöolosuhteissa, tarvitaan malleja, joiden avulla vikaantumisaajat voidaan muuntaa kuormitusta vastaaviksi.

Nopeutetun testauksen teoriassa oletetaan, että kuormituksen kasvattaminen ei muuta vikaantumismekanismeja, vaan ainoastaan nopeuttaa tapahtumia. Lineaarisen nopeutuksen tapauksessa jokaisen vikaantumisaajan tai jokaisen elinaikajakauman prosenttipisteen kertominen samalla vakiolla tuottaa elinaikajakauman projektion toisella kuormitustasolla. Olkoon t_U satunnainen vikaantumisaika tavanomaisissa käyttöolosuhteissa ja t_S aika, jolloin sama vikaantuminen olisi tapahtunut suuremmalla kuormituksella. Tällöin

$$t_U = AF \times t_S,$$

missä AF on nopeutuskerroin, joka suhteuttaa kahden eri kuormitustason vikaantumisaajat toisiinsa (Tobias & Trindade 1986). Olkoon F_U , f_U ja h_U kertymäfunktio, tiheysfunktio ja hasardifunktio tyypillisissä käyttöolosuhteissa, ja

F_S, f_S sekä h_S vastaavat funktiot kuormituksella S . Tällöin

$$\begin{aligned} F_U(t) &= F_S\left(\frac{t}{AF}\right) \\ f_U(t) &= \left(\frac{1}{AF}\right) f_S\left(\frac{t}{AF}\right) \\ h_U(t) &= \left(\frac{1}{AF}\right) h_S\left(\frac{t}{AF}\right). \end{aligned}$$

Nämä yhtälöt ovat voimassa yleisesti lineaarisen nopeutuksen tapauksessa, kun vikaantumismekanismit eivät muutu kuormituksen muuttuessa (Tobias & Trindade 1986).

Jos mallina käytetään lognormaalijakaumaa, saadaan kuormitustasolla S kertymäfunktiksi

$$F_S(t) = \Phi\left[\frac{\ln(t/T_{50(S)})}{\sigma_S}\right],$$

missä Φ on standardimuotoisen normaalijakauman kertymäfunktio ja $T_{50(S)}$ ja σ_S ovat lognormaalijakauman parametrit. Muunnoksella

$$F_U(t) = F_S\left(\frac{t}{AF}\right) = \Phi\frac{\ln[(t/AF)/T_{50(S)}]}{\sigma_S} = \Phi\frac{\ln[t/(AF \times T_{50(S)})]}{\sigma_S}$$

saadaan käyttöolosuhteiden kertymäfunktio, joka noudattaa myös lognormaalijakaumaa parametrein $\sigma_U = \sigma_S = \sigma$ ja $T_{50(U)} = AF \times T_{50(S)}$. Lineaarinen nopeutus ei muuta jakaumaa eikä jakauman muotoparametriä, ainoastaan skaalaparametri muuttuu nopeutuskertoimella kerrottaessa (Tobias & Trindade 1986).

5.6 Arrheniuksen malli

Arrheniuksen malli soveltuu tilanteisiin, joissa ollaan kiinnostuneita ainoastaan lämpötilasta kuormittavana tekijänä. Malli voidaan esittää muodossa

$$T_{50} = Ae^{\Delta H/kT},$$

missä A ja ΔH ovat tuntemattomia vakioita, k on Boltzmannin vakio ja T on lämpötila Kelvin-asteina. Boltzmannin vakion arvo on $1.380 \times 10^{-23} JK^{-1}$ tai $8.617 \times 10^{-5} eVK^{-1}$. Lämpötila saadaan muunnettua Kelvin-asteikolle lisäämällä Celsius-asteisiin 273.16. T_{50} :n sijasta voidaan käyttää mitä hyvänsä prosenttipistettä. Tällöin vakio A muuttuu, mutta muutos ei vaikuta nopeutuskertoimen arvoon. Malli perustuu Arrheniuksen lakiin, jonka mukaan yksinkertaisen kemiallisen reaktion nopeus riippuu lämpötilasta (Nelson 1990).

Nopeutuskerroin saadaan laskettua Arrheniuksen mallin avulla

$$AF = \frac{T_{50(1)}(\text{lämpötilassa } T_1)}{T_{50(2)}(\text{lämpötilassa } T_2)} = \frac{Ae^{\Delta H/kT_1}}{Ae^{\Delta H/kT_2}},$$

joten

$$AF = e^{(\Delta H/k)[(1/T_1)-(1/T_2)]}.$$

Tästä nähdään, että kahden lämpötilan välisen nopeutuskertoimen laskemiseksi riittää, kun tunnetaan ΔH . Vastaavasti, ΔH saadaan laskettua, jos nopeutuskerroin tunnetaan

$$\Delta H = k \left[\ln \left(\frac{T_{50(1)}}{T_{50(2)}} \right) \right] \left(\frac{1}{T_1} - \frac{1}{T_2} \right)^{-1}.$$

Tämän lausekkeen avulla saadaan estimoitua ΔH , kun käytössä on kahdessa eri lämpötilassa taltioituja vikaantumisaikoja.

5.7 Nopeutuskerroin paukkumaissiaineistolle

Paukkumaissiaineistolle laskettiin Arrheniuksen mallin avulla nopeutuskerroin. Laskennassa käytettiin 225- ja 250-asteisissa uuneissa mitattuja vikaantumisaikoja. Mediaanit olivat $T_{50(1)} \approx 89.4$ vikaantumisajoille 225 asteessa ja $T_{50(2)} \approx 79.2$ vikaantumisajoille 250 asteessa. Vakion ΔH arvoksi saatiin likimain 1.73×10^{-20} . Näillä tiedoilla pystyttiin laskemaan nopeutuskertoimen arvo $AF \approx 1.13$. Nopeutuskerroin voidaan tässä yhteydessä tulkita siten, että uunin lämpötilan pudottaminen 25 asteella kasvattaa vikaantumisaikat noin 1.13-kertaisiksi. Tätä voidaan hyödyntää esimerkiksi, kun halutaan arvioida vikaantumisaikoja 200-asteisessa uunissa. R-koodi nopeutuskertoimen laskemiseksi paukkumais-siaineistolle löytyy liiteosiosta.

6 Johtopäätökset

Aiheen lähestymistavasta johtuen ei tässä tutkielmassa esiinny varsinaisesti uusia tutkimustuloksia, ellei sellaisina voida pitää paukkumaissiesimerkkien tuloksia. Toisaalta tutkielman tarkoituksena ei ollut tuottaa uutta tietoa elintarviketeollisuudelle, vaan antaa luotettavuusanalyysistä kiinnostuneelle opiskelijalle perustietoja sekä konkreettinen esimerkkisovellus. Tulosten soveltaminen käytäntöön hyödyttäisi vain kaikkein intohimoisimpia paukkumaissin kotikäyttäjiä tai rasvattomien ja vähäsuolaisten pikkupurtavien ystäviä, sillä uunissa jyvät voi halutessaan valmistaa täysin ilman rasvaa.

Luotettavuusanalyysin kirjallisuus on yleensä kirjoitettu insinöörinäkökulmasta. Jos näkökulma on tilastotieteellinen, ei luotettavuudesta puhuta välttämättä mitään, vaan käytetään esimerkiksi termiä elinaika-analyysi. Vaikka terminologiassa on eroja, on asiasisältö pääpiirteissään sama. Erityisen haastavaa oli yrittää löytää suomenkielinen vastine kaikille luotettavuusanalyysin termeille. Termit pyrittiin varmentamaan useammasta lähteestä, lähinnä internetin avulla. Tässä onnistuttiin kohtalaisesti, mutta mukana on todennäköisesti vieläkin ilmauksia, joilla ei ole vakiintunutta asemaa suomenkielisessä luotettavuussanastossa.

Kirjallisuutta olisi voinut hyödyntää laajemminkin, mutta tässäkin aiheen käsittelytapa vaikutti toteutukseen: peruskäsitteet eivät muutu vuosien eivätkä välttämättä vuosikymmentenkään saatossa. Uusia menetelmiä kehitetään jatkuvasti ja vanhoja parannellaan. Aiheen raja-alue karsi kuitenkin lukuisia mielenkiintoisia luotettavuusanalyysin osa-alueita kokonaan tutkielman ulkopuolelle ja nopeutettu vikaantumisenkin käytiin läpi hyvin lyhyesti. Jakaumista mukaan piti ottaa alunperin eksponentti- ja Weibull-jakauma, mutta paukkumaissiaineiston myötä Weibull-jakauma vaihtui lognormaalijakaumaksi, jonka keralla oli luontevaa käsitellä normaalijakaumaa. Jakaumista eksponenttijakauma on saanut luultavasti parhaan ja yhtenäisimmän käsittelyn. Lognormaalijakauma on esitelty melko lyhyesti ja normaalijakauman osiossa suurimman uskottavuuden estimointi varastaa päähuomion. Näille ratkaisuille on kuitenkin olemassa myös perustelunsa. Suurimman uskottavuuden estimointi otettiin mukaan, jotta lukijalle kävisi selväksi minkälainen vaikutus sensuroinnilla on ja minkälaisia erityispiirteitä elinaikadata voi sisältää. Lognormaalijakauma käsiteltiin lyhyesti, koska lognormaalijakautunut data on helposti muunnettavissa normaalijakautuneeksi.

Jos paukkumaissiaineiston tutkimista haluaisi jatkaa, voisi yrittää selvittää virrankulutuksen kannalta optimaalisen paistolämpötilan. Ero kypsymisajois-

sa 225 asteessa ja 250 asteessa ei ollut enää kovin suuri, joten uunin virrankulutus on luultavasti suurempi 250 asteessa. Maissinjyvien mittaamista tai punnitusta ennen uuniin laittoa voisi kokeilla myös suuremmalla aineistolla ja eri lämpötiloissa. Tällöin kannattaisi arvioida myös lopputulosta eli auenneita maissinjyviä. Arvioinnissa voisi käyttää apuna parametrejä

$$\sigma = \frac{\text{aunneiden jyvien tilavuus (cm}^3\text{)}}{\text{otoksen massa ennen uuniin laittoa (g)}}$$

$$\pi = \frac{\text{aunneiden jyvien tilavuus (cm}^3\text{)}}{\text{aunneiden jyvien lkm}}$$

$$\omega = \frac{\text{aukeamattomien jyvien lkm}}{\text{kaikkien jyvien lkm}},$$

mutta tämä edellyttää selvästi suurempaa taloudellista ja ajallista panosta kuin mihin nyt oli mahdollisuuksia.

Lähdeluettelo

- Ebeling, Charles E. (1997), "An Introduction to Reliability and Maintainability Engineering", McGraw-Hill, Singapore.
- Hogg, Robert V. & Tanis, Elliot A. (1988), "Probability and Statistical Inference", Macmillan, New York.
- Lawless, Jerald F. (1982), "Statistical Models and Methods for Lifetime Data", John Wiley & Sons, New York.
- Leitch, Roger D. (1995), "Reliability Analysis for Engineers", Oxford University Press, New York.
- Liski, Erkki (2005), "Matemaattinen tilastotiede", Tampereen yliopisto, Opetusmoniste B 56.
- Mann, Nancy R., Schafer, Ray E. & Singpurwalla, Nozer D. (1974), "Methods for Statistical Analysis of Reliability and Life Data", John Wiley & Sons, New York.
- Nelson, Wayne (1990), "Accelerated Testing : statistical models, test plans, and data analyses", John Wiley & Sons, New York.
- Tobias, Paul A. & Trindade, David C. (1986), "Applied Reliability", Van Nostrand Reinhold, New York.

Liite 1: R-koodi eksponenttijakaumaesimerkille

lstlisting

```
# Generoidaan 100 havaintoa eksponenttijakaumasta vektoriin a
a <- rexp(100, rate=0.1)
# rate = lambda, jakauman keskiarvo on 1/lambda

# Järjestetään a:n arvot pienimmästä suurimpaan vektoriksi b
b <- sort(a)

# Piirretään b:n histogrammi
hist(b)

# lambda-hattu on esimerkissämme 100/sum(b) eli
# testattavien lkm jaettuna testitunneilla
c <- 100/sum(b)

# Muodostetaan aikajana
t <- c(1:60)

# Muodostetaan tiheysfunktio estimoidun lambda-hatun avulla
f <- c*exp(-c*t)

# Tiheysfunktion kuvaaja ja histogrammi samaan kuvaan.
# f kerrotaan 500:lla, koska tiheysfunktion kuvaajan
# alapuolelle jäävä ala on 1 ja histogrammin ala on
# n*(pylvään leveys) eli 100*5. Näin kuvaajat saadaan samaan
# mittakaavaan.
# Huom! Muokkaa tarvittaessa: hist(b) saattaa tuottaa myös
# pylväitä, joiden leveys on jotain muuta kuin 5.
plot(f*500, type="l", xlab="t", ylab="Vikaantumisten □ frekvenssi")
hist(b, add=TRUE)

# Khi-toiseen yhteensopivuustesti

# Muodostetaan osavälit siten että yksikään odotettu
# frekvenssi (vikaantumisten odotettu lkm) ei ole alle 5.
```

```

# Muokkaa tarvittaessa.
int1 <- b < 5
int2 <- (5 <= b) & (b < 10)
int3 <- (10 <= b) & (b < 15)
int4 <- (15 <= b) & (b < 25)
int5 <- 25 <= b

# Lasketaan vikaantumisten havaitut frekvenssit kullekin
# osavälille
O <- c(sum(int1),sum(int2),sum(int3),sum(int4),sum(int5))

# Lasketaan vikaantumisen todennäköisyys osavälillä
# kertymäfunktion avulla.
F <- 1-exp(-c*t)
P <- c(F[5],(F[10]-F[5]),(F[15]-F[10]),(F[25]-F[15]),(1-F[25]))

# Lasketaan vikaantumisten odotettu lkm kertomalla
# vikaantumisten todennäköisyydet otoskoolla eli 100:lla.
E <- 100*P

# Lasketaan khi-toiseen -testisuure
Khi <- sum((O - E)^2/E)

# Khi toiseen -jakauman 90. ja 95. prosenttipiste, joihin
# yhteensopivuustestin testisuureta voi verrata. Vapausasteet
# saadaan vähentämällä osavälien määrästä estimoitujen
# parametrien lukumäärä ja yksi. Koska estimoituja parametrejä
# on nyt yksi, lambda, saadaan vapausasteiksi 5-1-1 = 3.
qchisq(c(0.90,0.95), 3, ncp=0, lower.tail=TRUE, log.p=FALSE)

# Luottamusväli lambdalle (Type 2 censored data)

# 95 prosentin lv:n alarajaa varten khi toiseen -jakauman 2.5
# %:n prosenttipiste
# (df=2r, missä r on vikaantumisten lkm eli 100)
lt <- qchisq(0.025, 200, ncp=0, lower.tail=TRUE, log.p=FALSE)

# Khi toiseen -jakauman 97.5 %:n prosenttipiste
ut <- qchisq(0.025, 200, ncp=0, lower.tail=FALSE, log.p=FALSE)

# lv:n ylä- ja alarajoja varten tarvittavat kertoimet
# k(r; alpha/2) ja k(r; 1-alpha/2)
klt <- lt/200
kut <- ut/200

# lambdan 95 %:n lv

```

```
lv <- c(klt*c, kut*c)
```

```
# Luottamusväli lambdaalle (Type 1 censored data)  
# Muuten kuten edellä, mutta khi toiseen -jakauman  
# yläprosenttipistettä laskettaessa vapausasteet ovat 2(r+1).  
# Voidaan laskea vain, jos vikaantuneet yksiköt on korvattu  
# välittömästi uusilla.  
lt <- qchisq(0.025, 200, ncp=0, lower.tail=TRUE, log.p=FALSE)  
ut <- qchisq(0.025, 202, ncp=0, lower.tail=FALSE, log.p=FALSE)  
# Huom! muutos df=202  
klt <- lt/200  
kut <- ut/200  
lv <- c(klt*c, kut*c)
```

Liite 2: R-koodi lognormaalijakaumaesimerkille

lstlisting

```
# Syötetään R:lle paukkumaissidata vektoriksi pop  
# (uunin lämpötila 200 astetta).  
  
a <- c(106, 126, 133, 138, 143, 147, 150, 152, 172, 182)  
b <- c(135, 148, 149, 154, 171, 174, 182, 208, 266)  
c <- c(103, 128, 137, 143, 148, 154, 180, 182, 191, 198)  
d <- c(80, 124, 127, 134, 143, 152, 155, 170, 183)  
e <- c(119, 127, 130, 152, 159, 166, 173, 179, 212)  
f <- c(133, 136, 150, 159, 160, 165, 195, 211, 228, 238)  
g <- c(126, 130, 133, 140, 152, 154, 182, 185, 191, 196)  
h <- c(100, 129, 158, 171, 184, 190, 192, 210, 217, 259)  
i <- c(127, 135, 142, 143, 147, 150, 155, 158, 161, 168)  
j <- c(113, 119, 130, 137, 142, 146, 164, 170, 200, 219)  
k <- c(94, 115, 121)  
  
pop <- c(a, b, c, d, e, f, g, h, i, j, k)  
  
# Järjestetään arvot pienimmästä suurimpaan vektoriksi pop2  
  
pop2 <- sort(pop)  
  
# Otetaan arvoista logaritmit, jolloin (oletuksen mukaan)  
# vikaantumisaajat noudattavat normaalijakaumaa.  
  
lpop <- log(pop2)  
  
# Vikaantumisaikojen logaritmien histogrammi  
  
hist(lpop)  
  
# Jos kaikkien testiyksiköiden vikaantumisaajat ovat tiedossa,  
# saadaan normaalijakauman parametrit helposti estimoitua.  
# Myyn estimaatti:
```

```

myhat <- sum(lpop)/sum(lpop/lpop) # tai myhat <- mean(lpop)

# Sigman estimaatti:

shat <- sqrt(sum((lpop-mean(lpop))^2/(sum(lpop/lpop)-1)))

# Noudattavatko vikaantumisaikojen logaritmit normaalijakaumaa?
# Shapiro-Wilk -testi (vain sensuroimattomalle datalle):

shapiro.test(lpop)

# t-jakauman 97.5 %:n prosenttipiste, vapausasteilla n - 1 = 99

t <- qt(.975, df = 99)

# myyn luottamusvälin ylä- ja alaraja 95 %:n luottamustasolla:

lvala <- mean(lpop) - t*(shat/sqrt(99))
lvyla <- mean(lpop) + t*(shat/sqrt(99))

# Khi2-jakauman 2.5 %:n ja 97.5 %:n prosenttipisteet,
# vapausasteilla n - 1 = 99

chi <- qchisq(c(0.025,0.975), df = 99)

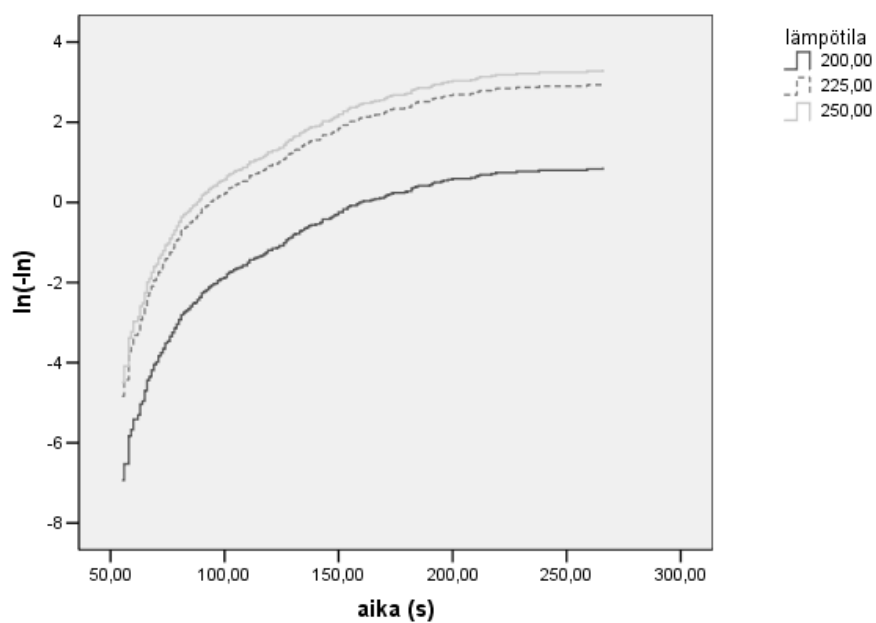
# sigma2:n luottamusvälin ylä- ja alaraja 95 %:n
# luottamustasolla:

cil <- (99*shat^2)/chi[2]
ciu <- (99*shat^2)/chi[1]

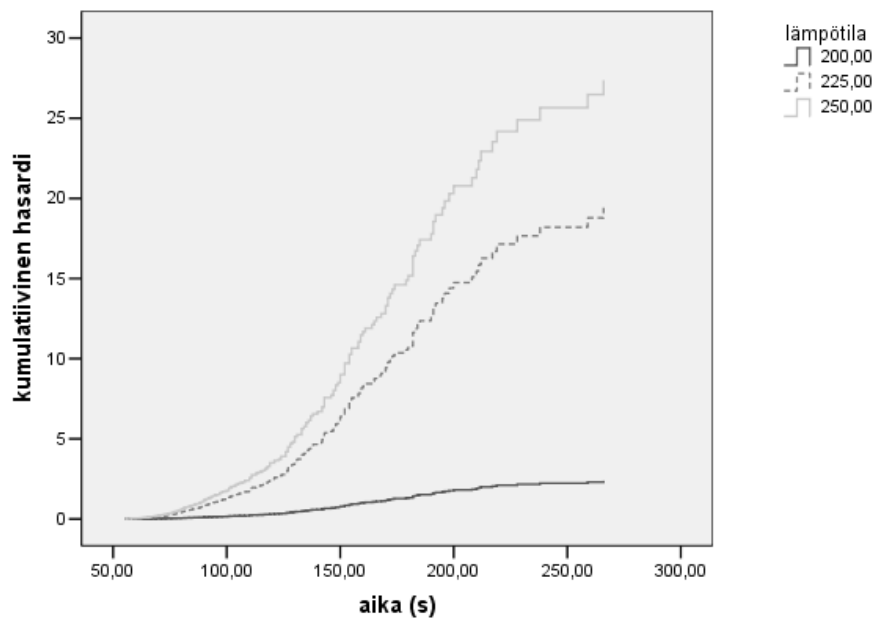
```

Liite 3: Coxin regression kuvaajia

$\ln(-\ln R(t))$ -funktion luokkakohtaiset kuvaajat paukkumaissiaineistolle



Kumulatiivisen hasardifunktion luokkakohtaiset kuvaajat paukkumaissiaineistolle



Liite 4: R-koodi nopeutus kertoimen laskemiseksi

lstlisting

```
# Määritellään paukkumaissidatan nopeutuskerroin  
# Arrheniuksen mallin avulla.  
  
# Lognormaalijakauman mediaani, T(50), 225 ja 250  
# Celsius-asteessa.  
Med1 <- exp(4.492993)  
Med2 <- exp(4.372417)  
  
# Boltzmannin vakio  
Boltz <- 1.380*10^(-23)  
  
# Lämpötilat Kelvin-asteina  
T1 <- 225 + 273.16  
T2 <- 250 + 273.16  
  
# Lasketaan deltaH  
deltaH <- Boltz*(log(Med1/Med2))*((1/T1)-(1/T2))^(-1)  
  
# Nyt voidaan laskea nopeutuskerroin  
AF <- exp((deltaH/Boltz)*((1/T1)-(1/T2)))
```