Differentiaalientropia (MATA2500)

Sami Yrjänheikki

3. joulukuuta 2019

Tiivistelmä

Tässä esseessä tarkastellaan differentiaalientropiaa, sen määritelmää ja ominaisuuksia sekä yhteyttä Shannonin entropiaan ja Kullback–Leibler divergenssiin. Lisäksi tarkastellaan differentiaalientropian alkuperäisen määritelmän ongelmia ja lasketaan muutamien jakaumien entropioita. Keskeisiä todistuksia ovat Gibbsin epäyhtälö ja normaalijakauman entropian maksimointi. Lopuksi tehdään pikakatsaus differentiaalientropian sovelluksiin.

1 Differentiaalientropian määritelmä

Todetaan heti alkuun, että tässä esseessä käytetään luonnollista logaritmia kaksikantaisen sijaan. Tämä siksi, että matemaattisesti luonnollinen logaritmi on helpompi käsitellä. Logaritmin kannanvaihdosta $\log_2(x) = \log(x)/\log 2$ nähdään, että logaritmin kantaluku vaikuttaa laskuihin vain jonkin vakion ($1/\log 2 \approx 1.44$) verran. Kyse on siis eri yksiköistä: kaksikantaisella logaritmilla lasketun entropian yksikkö on bitti, kun taas luonnollisella logaritmilla yksikkö on *nat*.

Aloitetaan sitten varsinainen differentiaalientropian tarkastelu määrittelemällä aluksi *Shannonin entropian* eli diskreetti entropia [1].

Määritelmä 1.1 (Shannonin entropia)**.** Olkoon (Ω, F, P) äärellinen todennäköisyysavaruus ja $X : \Omega \to \mathbb{R}$ satunnaismuuttuja. Määritellään Shannonin entropiaksi

$$H(X) = -\sum_{x\in\Omega} P(x)\log(P(x)).$$

Olisi kuitenkin melko luontevaa laajentaa entropian käsite myös äärettömiin todennäköisyysavaruuksiin, joissa voi esiintyy jatkuvia todennäköisyysjakaumia. Heuristisesti äärellinen summa korvaantuisi tällöin integraalilla ja aikoinaan Shannon määrittelikin *differentiaalientropian käsitteen* juuri näin [2, s. 243].

Määritelmä 1.2 (Differentiaalientropia)**.** Olkoon (Ω, F, P) todennäköisyysavaruus ja $X : \Omega \to \mathbb{R}$ satunnaismuuttuja. Määritellään differentiaalientropiaksi

$$H(X) = -\int_\Omega P(x)\log(P(x))\,\mathrm{d}x.$$

Eräs differentiaalientropian käsitteen hyödyistä on se, että entropian laskemiseen saadaan käyttöön integraalilaskennan koneisto. Shannonin entropiassa olevan äärellisen summan laskeminen symbolisesti voi monesti osoittautua vaikeaksi.

Esimerkki 1.3. Olkoon (Ω, F, P) todennäköisyysavaruus, jossa $\Omega = \{x_1, \ldots, x_n\} \subset \mathbb{R}$ on jaoteltu tasavälein siten, että $x_{i+1} - x_i = \Delta x \in \mathbb{R}$ kaikilla $i \in \{1, \ldots, n\}$. Määritellään sitten jatkuva kuvaus $\tilde{P}(x) = P(x_i)$ kun $x \in [x_i, x_{i+1}]$. Tämän uuden todennäköisyysfunktion $\tilde{P}$ differentiaalientropiaksi tulee määritelmän ja integraalin lineaarisuuden nojalla

$$\begin{aligned}
H(\tilde{P}) &= -\int_\Omega \tilde{P}(x)\log(\tilde{P}(x))\,\mathrm{d}x = -\sum_{i=1}^{n}\int_{x_i}^{x_{i+1}} P(x_i)\log((P(x_i))\,\mathrm{d}x \\
&= -\sum_{i=1}^{n} P(x_i)\log(P(x_i))\int_{x_i}^{x_i+\Delta x}\mathrm{d}x = -\sum_{x\in\Omega} P(x)\log(P(x))\Delta x \\
&= H(P)\Delta x.
\end{aligned}$$

Havaitaan siis, että tasavälisen porrasjakauman differentiaalientropia vastaa diskreettiä Shannonin entropiaa silloin, kun $\Delta x = 1$, kuten voitaisiinkin odottaa.

2 Todennäköisyysjakaumien entropioita

Monet tärkeät ja sovelluksissa paljon käytetyt todennäköisyysjakaumat ovat luonteeltaan jatkuvia, eikä niihin siten voida liittää Shannonin entropiaa. Sen sijaan niille voidaan laskea differentiaalientropia. Tarkastellaan seuraavaksi millaisia differentiaalientropioita tulee yleisimmille jatkuville todennäköisyysjakaumille.

Tasajakauma Tiheysfunktion $u : [a, b] \to \mathbb{R}$,

$$u(x) = \frac{1}{b-a}$$

entropiaksi saadaan suoraan määritelmästä laskemalla

$$\begin{aligned}
H(u) &= -\int_a^b u(x)\log(u(x))\,\mathrm{d}x \\
&= \int_a^b \frac{\log(b-a)}{b-a}\,\mathrm{d}x \\
&= \frac{\log(b-a)}{b-a}(b-a) \\
&= \log(b-a).
\end{aligned}$$

Normaalijakauma Normaalijakauman tiheysfunktion $p : \mathbb{R} \to \mathbb{R}$,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

entropiaksi voidaan laskea

$$\begin{aligned}
H(p) &= -\int_{\mathbb{R}} p(x)\log(p(x))\,\mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}}\left(-\int_{\mathbb{R}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\mathrm{d}x + \frac{1}{2\sigma^2}\int_{\mathbb{R}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)(x-\mu)^2\,\mathrm{d}x\right) \\
&= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2} \\
&= \log(\sigma\sqrt{2\pi e}).
\end{aligned}$$

Eksponenttijakauma Eksponenttijakauman tiheysfunktiolle $e : [0, \infty[\to \mathbb{R}$,

$$e(x) = \lambda\exp(-\lambda x)$$

saadaan entropiaksi

$$
\begin{aligned}
H(e) &= -\int_0^\infty e(x)\log(e(x))\,\mathrm{d}x \\
&= -\lambda\int_0^\infty \exp(-\lambda x)(\log(\lambda)-\lambda x)\,\mathrm{d}x \\
&= -\lambda\left(\log(\lambda)\int_0^\infty \exp(-\lambda x)\,\mathrm{d}x - \int_0^\infty \frac{\lambda x}{\exp(\lambda x)}\,\mathrm{d}x\right) \\
&= -\lambda\left(\frac{\log(\lambda)}{\lambda}-\frac{1}{\lambda}\right) \\
&= 1-\log(\lambda).
\end{aligned}
$$

Maxwell–Boltzmann-jakauma Statistisessa fysiikassa keskeisen Maxwell–Boltzmann-jakauman tiheysfunktion $f_{\mathrm{MB}} : [0,\infty[\, \to \mathbb{R}$,

$$
f_{\mathrm{MB}}(x) = \frac{1}{a^3}\sqrt{\frac{2}{\pi}}x^2\exp\left(-\frac{x^2}{2a^2}\right)
$$

entropian laskemiseen tarvittavia integraaleja ei voida esittää alkeisfunktioiden avulla, mutta osoittautuu, että

$$
\begin{aligned}
H(f_{\mathrm{MB}}) &= -\int_0^\infty f_{\mathrm{MB}}(x)\log(f_{\mathrm{MB}}(x))\,\mathrm{d}x \\
&= \log(a\sqrt{2\pi})+\gamma-\frac{1}{2}
\end{aligned}
$$

missä $\gamma \approx 0.577$ on Eulerin–Mascheronin vakio [2, s. 662].

Kerätään vielä yhteen saadut entropiat eri jakaumille seuraavaan lauseeseen.

Lause 2.1. *Jatkuvien jakaumien tiheysfunktioiden entropioille pätee*

$$
\begin{aligned}
H &= \log(b-a) && \textit{tasajakauma } \mathcal{U}(a,b), \\
H &= \log(\sigma\sqrt{2\pi e}) && \textit{normaalijakauma } \mathcal{N}(\mu,\sigma^2), \\
H &= 1-\log(\lambda) && \textit{eksponenttijakauma } \mathrm{Exp}(\lambda), \\
H &= \log(a\sqrt{2\pi})+\gamma-\frac{1}{2} && \textit{Maxwell–Boltzmann-jakauma } \mathrm{MB}(a).
\end{aligned}
$$

Todistus. Laskut esitetty aiemmin tässä osiossa. □

3 Gibbsin epäyhtälö

Eräs tärkeä differentiaalientropiaan liittyvä tulos on niin kutsuttu Gibbsin epäyhtälö, sillä sen avulla voidaan osoittaa useita entropiaan liittyviä ominaisuuksia [2]. Annetaan sitä ennen kuitenkin logaritmille eräs yläraja ja määritellään *Kullback–Leibler divergenssin*[1] käsite, joka on hyödyllinen tapa esittää Gibbsin epäyhtälö [2, s. 251]. Kullback–Leibler divergenssi antaa myös tavan vertailla kahta todennäköisyysjakaumaa ja niiden entropioita, mutta tämä näkökulma jää vähälle huomiolle.

Lemma 3.1. *Logaritmifunktiolle* $\log :]0, \infty[\to \mathbb{R}$ *pätee* $\log(x) \leq x - 1$ *kaikilla* $x > 0$.

Todistus. Tarkastellaan funktiota $f :]0, \infty[\to \mathbb{R}$, $f(x) = x - 1 - \log(x)$. Sille pätee

$$f'(x) = 1 - \frac{1}{x} = 0 \iff x = 1$$

ja $f''(1) = 1 > 0$, joten f:llä on minimi pisteessä $(1, f(1))$. Toisaalta $f(1) = 0$ eli

$$f(x) = x - 1 - \log(x) \geq f(1) = 0 \iff \log(x) \leq x - 1$$

kaikilla $x > 0$, kuten haluttiinkin. □

Määritelmä 3.2 (Kullback–Leibler divergenssi)**.** Olkoon P ja Q samassa joukossa $\Omega \subset \mathbb{R}$ määriteltyjä todennäköisyysjakaumia. Määritellään Kullback–Leibler divergenssiksi *(KL-divergenssi)*

$$D(P \| Q) = \int_\Omega P(x) \log\left(\frac{P(x)}{Q(x)}\right) \mathrm{d}x.$$

Näiden työkalujen avulla voimme nyt muotoilla ja todistaa Gibbsin epäyhtälön.

Lause 3.3 (Gibbsin epäyhtälö)**.** *KL-divergenssi on aina ei-negatiivinen eli*

$$D(P \| Q) = \int_\Omega P(x) \log\left(\frac{P(x)}{Q(x)}\right) \mathrm{d}x \geq 0$$

millä tahansa kahdella samassa joukossa $\Omega \subset \mathbb{R}$ *määritellylle todennäköisyysjakaumalle* P *ja* Q.

Todistus. Jos $P(x) \neq 0$ ja $Q(x) \neq 0$ kaikilla $x \in \Omega$, niin Lemman 3.1 nojalla

$$P(x) \log\left(\frac{Q(x)}{P(x)}\right) \leq P(x) \left(\frac{Q(x)}{P(x)} - 1\right) = Q(x) - P(x),$$

joten logaritmin laskusäännöillä

$$P(x) \log\left(\frac{P(x)}{Q(x)}\right) \geq P(x) - Q(x).$$

Tällöin integraalin monotonisuuden nojalla

[1] Tunnetaan myös nimellä *suhteellinen entropia (relative entropy)*.

$$\int_\Omega P(x)\log\left(\frac{P(x)}{Q(x)}\right)\,\mathrm{d}x \geq \int_\Omega \left(P(x)-Q(x)\right)\,\mathrm{d}x.$$

Kolmogorovin aksioomien perusteella

$$\int_\Omega P(x)\,\mathrm{d}x = \int_\Omega Q(x)\,\mathrm{d}x = 1,$$

joten integraalin lineaarisuuden nojalla

$$\int_\Omega \left(P(x)-Q(x)\right)\,\mathrm{d}x = \int_\Omega P(x)\,\mathrm{d}x - \int_\Omega Q(x)\,\mathrm{d}x = 1-1=0.$$

Jos $Q(x)=0$ jollakin $x\in\Omega$, jatkuvuuden[2] perusteella voidaan asettaa

$$P(x)\log\left(\frac{P(x)}{Q(x)}\right)=\infty,$$

jolloin $D(P\|Q)$ kasvaa rajatta, mutta erityisesti $D(P\|Q)\geq 0$. Jos puolestaan $P(x)=0$ jollakin $x\in\Omega$ ja $Q(x)\neq 0$ kaikilla $x\in\Omega$, voidaan jälleen jatkuvuuden[3] nojalla asettaa

$$P(x)\log\left(\frac{P(x)}{Q(x)}\right)=0,$$

joten erityisesti $D(P\|Q)\geq 0$. Edeltävistä nähdään myös, että jos $P(x)=0$ kaikilla $x\in\Omega$, niin $D(P\|Q)=0$. Jos taas $P(x)=Q(x)$ kaikilla $x\in\Omega$, niin

$$D(P\|Q)=\int_\Omega P(x)\underbrace{\log\left(\frac{P(x)}{Q(x)}\right)}_{=0}\mathrm{d}x=0.$$

□

[2] $P(x)\log(P(x)/Q(x))\to\infty$ kun $Q(x)\to 0^+$ logaritmin monotonisuuden nojalla.
[3] $P(x)\log(P(x)/Q(x))\to 0$ kun $P(x)\to 0^+$ l'Hôpitalin säännön nojalla.

4 Entropian maksimointi

Normaalijakauman monista mielenkiintoisista ominaisuuksista keskitymme seuraavaksi differentiaalientropian maksimointiin. Osoittautuu, että kaikista todennäköisyysjakaumista nimenomaan normaalijakauman entropia on suurin[4], kuten seuraavaksi nähdään. Määritellään vielä sitä ennen normaalijakauman tiheysfunktio täsmällisesti.

Määritelmä 4.1 (Normaalijakauman tiheysfunktio)**.** Normaalijakauman $\mathcal{N}(\mu, \sigma^2)$ todennäköisyystiheysfunktio on $p : \mathbb{R} \to [0, \infty[$,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Lause 4.2 (Entropian maksimointi)**.** *Olkoon $p : \mathbb{R} \to [0, \infty[$ normaalijakauman $\mathcal{N}(\mu, \sigma^2)$ tiheysfunktio ja $f : \mathbb{R} \to [0, \infty[$ tiheysfunktio jollekin mielivaltaiselle todennäköisyysjakaumalle, jolla on sama varianssi σ^2. Tällöin $H(p) \geq H(f)$.*

Todistus. Logaritmin laskusäännöistä seuraa, että

$$D(f\|p) = \int_{\mathbb{R}} f(x) \log\left(\frac{f(x)}{p(x)}\right) \mathrm{d}x = -H(f) - \int_{\mathbb{R}} f(x) \log(p(x))\, \mathrm{d}x.$$

Gibbsin epäyhtälöstä saadaan, että

$$D(f\|p) = -H(f) - \int_{\mathbb{R}} f(x) \log(p(x))\, \mathrm{d}x \geq 0,$$

joten

$$-\int_{\mathbb{R}} f(x) \log(p(x))\, \mathrm{d}x \geq H(f).$$

Riittää siis osoittaa, että

$$\int_{\mathbb{R}} f(x) \log(p(x))\, \mathrm{d}x = -H(p).$$

Suoralla laskulla saadaan, että

$$\begin{aligned}
\int_{\mathbb{R}} f(x) \log(p(x))\, \mathrm{d}x &= \int_{\mathbb{R}} f(x) \log\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) \mathrm{d}x \\
&= \int_{\mathbb{R}} f(x) \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) \mathrm{d}x + \int_{\mathbb{R}} f(x) \log\left(\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) \mathrm{d}x \\
&= -\frac{1}{2}\log(2\pi\sigma^2) \int_{\mathbb{R}} f(x)\, \mathrm{d}x - \frac{1}{2\sigma^2} \int_{\mathbb{R}} f(x)(x-\mu)^2\, \mathrm{d}x.
\end{aligned}$$

Kolmogorovin aksioomien perusteella

$$\int_{\mathbb{R}} f(x)\, \mathrm{d}x = 1.$$

[4]Kiinnitetylle varianssille σ^2, kuten käy ilmi Lauseen 4.2 oletuksista.

Varianssin σ^2 määritelmästä[5] seuraa puolestaan, että

$$\int_{\mathbb{R}} f(x)(x-\mu)^2 \, \mathrm{d}x = \sigma^2,$$

joten

$$\int_{\mathbb{R}} f(x) \log(p(x)) \, \mathrm{d}x = -\frac{1}{2} \left(\log(2\pi\sigma^2) + 1 \right).$$

Koska saadun integraalin arvo ei riipu tiheysfunktiosta f, pätee tällöin myös

$$H(p) = -\int_{\mathbb{R}} p(x) \log(p(x)) \, \mathrm{d}x = \frac{1}{2} \left(\log(2\pi\sigma^2) + 1 \right),$$

joten yhdistämällä nämä tulokset saadaan

$$\int_{\mathbb{R}} f(x) \log(p(x)) \, \mathrm{d}x = -H(p),$$

kuten haluttiinkin. $\square$

[5] Varianssin yleisestä määritelmästä $\sigma^2 := \mathbb{E}[(X-\mu)^2]$ seuraa jatkuvan jakauman tapauksessa muoto $\sigma^2 = \int_{\Omega} f(x)(x-\mu)^2 \, \mathrm{d}x$.

5 Differentiaalientropian kritiikkiä

Shannon määritteli aikoinaan differentiaalientropian käsitteen korvaamalla äärellisen summan suoraan integraalilla [4]. Vaikka tämä on heuristisesti järkevän kuuloista *("integraali on ääretön summa")*, osoittautuu, että vaihto integraaliin ei olekaan niin ongelmatonta.

Entropian tarkoitus on kuvata jonkin jakauman keskimääräistä (odotusarvollista) tapahtuman siirtämää informaatiota. On siis melko selvää, että minkään jakauman entropia ei voisi olla negatiivinen, sillä tämä tarkoittaisi tilannetta, jossa keskimäärin tapahtumat *kasvattaisivat* epävarmuutta. Shannonin entropia toteuttaakin tämän melko luonnollisen ehdon [2, s. 15], mutta differentiaalientropian kohdalla tämä ehto ei toteudu.

Esimerkki 5.1 (Negatiivinen entropia)**.** Olkoon $f : \left[0, \frac{1}{e}\right] \to \mathbb{R}$ tasajakauman $\mathcal{U}(0, \frac{1}{e})$ tiheysfunktio. Tällöin Lauseen 2.1 mukaan

$$H(f) = \log\left(\frac{1}{e} - 0\right) = -1 < 0.$$

Toisaalta on myös mahdollista, että differentiaalientropian sisältämä integraali ei suppene ollenkaan vaan hajaantuu äärettömyyteen.

Esimerkki 5.2 (Ääretön entropia)**.** Määritellään todennäköisyysjakauman tiheysfunktioksi [3] $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \begin{cases} \dfrac{\log 2}{x \log(x)^2}, & x > 2 \\ 0, & x \leq 2. \end{cases}$$

Selvästi f on tiheysfunktio, sillä $f(x) \geq 0 \;\; \forall x \in \mathbb{R}$ ja

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_2^\infty \frac{\log 2}{x \log(x)^2}\,\mathrm{d}x = \lim_{x\to\infty}\left(-\frac{\log 2}{\log(x)}\right) + 1 = 1.$$

Kuitenkin

$$\begin{aligned} \int_{\mathbb{R}} f(x) \log\left(f(x)\right)\,\mathrm{d}x &= \int_{\mathbb{R}} \frac{\log 2}{x\log(x)^2} \log\left(\frac{\log 2}{x \log(x)^2}\right)\,\mathrm{d}x \\ &= \lim_{x\to\infty} \log 2 \left(\frac{2}{\log(x)} - \frac{\log\left(\frac{\log 2}{x\log^2(x)}\right)}{\log(x)} - \log(\log(x))\right) + \underbrace{\alpha}_{\in\mathbb{R}} \\ &= -\infty, \end{aligned}$$

joten $H(f) = \infty$.

Differentiaalientropialta puuttuu myös muita ominaisuuksia, joita diskreetillä entropialla on, kuten invarianssi muuttujanvaihdossa ja skaalauksessa [4]. Näistä ongelmista johtuen differentiaalientropiaa ei voi ajatella absoluuttisena epävarmuuden mittarina vaan korkeintaan suhteellisena. E. T. Jaynes onkin esittänyt vaihtoehtoista määritelmää differentiaalientropialle, joka nojaa vahvemmalle matemaattiselle pohjalle [5].

Määritelmä 5.3 (Vaihtoehtoinen differentiaalientropian määritelmä)**.** Olkoon (Ω, F, P) todennäköisyysavaruus ja $X : \Omega \to \mathbb{R}$ satunnaismuuttuja. Asetetaan

$$H(X) = -\int_{\Omega} P(x) \log \left(\frac{P(x)}{m(x)} \right) \mathrm{d}x = -D(P \| m),$$

missä

$$\int_a^b m(x)\,\mathrm{d}x = \lim_{N \to \infty} \frac{|\{x \in \Omega : a < x < b\}|}{N},$$

missä $|\Omega| = N$.

Määritelmä 5.3 antaa siis differentiaalientropian variantin, jossa Shannonin versioon on lisätty tekijä $m(x)$. Tähän vaihtoehtoiseen versioon ei kuitenkaan tutustuta sen tarkemmin, joten tekijän $m(x)$ hieman mystistä määritelmää[6] ei tarkastella lähemmin.

6 Sovelluksia

Vaikka edellisessä osiossa nähtiin, että Shannonin differentiaalientropian määritelmä ei ole kovinkaan vankalla matemaattisella pohjalla, sillä on silti hyötynsä. Eräs jo esille tullut hyöty on eri jakaumien entropioiden laskeminen ja vertailu. Laskemista helpoittaa erityisesti se, ettei Määritelmän 5.3 tekijää $m(x)$ tarvitse huomioida. Differentiaalientropia antaa myös kätevän koneiston normaalijakauman entropian maksimoinnin todistukselle. Samalla koneistolla voisikin todistaa, miten muut jakaumat maksimoivat entropian toisenlaisilla ehdoilla. Esimerkiksi kysymykseen “Mikä jatkuva jakauma maksimoi entropian välillä $]0, \infty[$ kiinnitetyllä odotusarvolla $\mu = 1/\lambda$?” saatu vastaus “Eksponenttijakauma $\mathrm{Exp}(\lambda)$” voidaan osoittaa samalla differentiaalientropian koneistolla [4].

Samalla luentosarjalla missä Jaynes esitti Määritelmän 5.3, hän sovelsi differentiaalientropiaa statistiseen fysiikkaan pyrkien perustelemaan yleisesti käytettyjen jakaumien soveltuvuutta [5]. Lisää esimerkkejä nimenomaan differentiaalientropian sovelluksista löytyy esimerkiksi materiaalitieteistä [6], elektroniikasta [7] ja tietojenkäsittelytieteistä [8].

[6] Jaynesin differentiaalientropian idea on lähteä liikkeelle diskreetistä joukosta ja tarkastella tällaisen joukon tiheyttä rajalla $N \to \infty$, missä N on joukon alkioiden lukumäärä.

Viitteet

[1] Claude E. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal, 1948.

[2] Thomas M. Cover, Joy A. Thomas, *Elements of Information Theory*, Wiley, 2006

[3] syeh_106, *Is differential entropy always less than infinity?*, Cross Validated, 2015 (luettu 27.11.2019)
https://stats.stackexchange.com/q/156252

[4] Charles Marsh, *Introduction to Continuous Entropy*, Department of Computer Science, Princeton University,
https://www.crmarsh.com/static/pdf/Charles_Marsh_Continuous_Entropy.pdf

[5] E. T. Jaynes, *Information Theory and Statistical Mechanics*, Brandeis University Summer Institute Lectures in Theoretical Physics, 1962,
https://bayes.wustl.edu/etj/articles/brandeis.pdf

[6] Mu-Hang Zhang, Xiao-Hong Shen, Lei He, Ke-Shi Zhang, *Application of Differential Entropy in Characterizing the Deformation Inhomogeneity and Life Prediction of Low-Cycle Fatigue of Metals*, Materials, 2018

[7] Mervat Mahdy, Dina S Eltelbany *Application of Entropy Measures to a Failure Times of Electrical Components Models*, Pakistan Journal of Statistics and Operation Research, 2017

[8] Aapo Hyvärinen, *New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit*, Advances in Neural Information Processing Systems, 1998
https://www.cs.helsinki.fi/u/ahyvarin/papers/NIPS97.pdf