A series of maximum entropy upper bounds of the differential entropy

https://arxiv.org/abs/1612.02954
https://www.lix.polytechnique.fr/~nielsen/MEUB/

Frank Nielsen[1,2] Richard Nock[3,4,5]
Frank.Nielsen@acm.org

[1]École Polytechnique, France [2]Sony Computer Science Laboratories, Japan
[3]Data61, Australia [4]ANU, Australia [5]The University of Sydney, Australia

January 2017

Shannon's differential entropy

$X \sim p(x)$: continuous random variable, support
$\mathcal{X} = \{x \in \mathbb{R} \ : \ p(x) > 0\}$
Shannon's entropy quantifies amount of uncertainty [2]:

$$\boxed{H(X) = \int_{\mathcal{X}} p(x) \log \frac{1}{p(x)} \mathrm{d}x = -\int_{\mathcal{X}} p(x) \log p(x) \mathrm{d}x} \qquad (1)$$

logarithm: basis 2 (unit in *bits*), basis e (*nats*).
Differential entropy is strictly concave and:

- May be negative: $X \sim N(\mu, \sigma)$, $H(X) = \frac{1}{2}\log(2\pi e \sigma^2) < 0$ when $\sigma < \frac{1}{\sqrt{2\pi e}}$
- May be infinite (unbounded): $X \sim p(x)$ with $p(x) = \frac{\log(2)}{x \log^2 x}$ for $x > 2$ (with support $\mathcal{X} = (2, \infty)$)
- Closed forms [7, 9] for many distribution families, but the *differential entropy of mixtures* usually does not admit closed-form expressions [10, 6]

Maximum Entropy Principle (MaxEnt)

Jaynes' MaxEnt distribution principle [4, 5] (1957):
Infer a distribution given several moment constraints.
Constrained optimization problem:

$$\max_p H(p) \;:\; E[t_i(X)] = \eta_i, \quad i \in [D] = \{1, \ldots, D\}. \tag{2}$$

- When an iid sample set $\{x_1, \ldots, x_s\}$ is given, we may choose, for example, the raw geometric *sample moments* $\eta_i = \frac{1}{s}\sum_{j=1}^{s} x_j^i$ for setting up the constraint $E[X^i] = \eta_i$ (ie., taking $t_i(X) = X^i$ in Eq. 2).
- The distribution $p(x)$ maximizing the entropy under those moment constraints is unique and termed the *MaxEnt distribution*. The constrained optimization of Eq. 2 is solved by means of Lagrangian multipliers [8, 2].

MaxEnt and exponential families

MaxEnt distribution $p(x)$ belongs to a *parametric family* of distributions called an *exponential family* [1, 8, 3].

Canonical probability density function of an exponential family (EF):

$$\boxed{p(x;\theta) = \exp\left(\langle\theta, t(x)\rangle - F(\theta)\right)} \tag{3}$$

$\langle a, b\rangle = a^\top b$: scalar product
$\theta \in \Theta$: natural parameter
Θ: natural parameter space
$t(x)$: sufficient statistics
$F(\theta) = \log \int p(x;\theta)\mathrm{d}x$: log-normalizer [1]

Dual parameterizations of exponential families

A distribution $p(x;\theta)$ of an exponential family can be parameterized equivalently either using the

- natural coordinate system θ,
- expectation coordinate system $\eta = E_{p(x;\theta)}[t(x)]$ (also called moment coordinate system)

The two coordinate systems are linked by the Legendre transformation:

$$F^*(\eta) = \sup_{\theta}\{\langle \eta, \theta \rangle - F(\theta)\}$$

$$\eta = \nabla F(\theta), \quad \theta = \nabla F^*(\eta)$$

In practice, when $F(\theta)$ is not available in closed-forms, conversion $\theta \leftrightarrow \eta$ is approximated numerically [8].

Differential entropy of exponential families

Closed-form when the dual Legendre convex conjugate function is in closed-form:

$$\boxed{H(p(x;\theta)) = -F^*(\eta(\theta))}$$

More general form when allowing an auxiliary carrier measure term [9]

Strategy to get MaxEnt Upper Bounds (MEUBs)

Rationale: *Any* other distribution with density $p'(x)$ different from the MaxEnt distribution $p(x)$ and satisfying all the D moment constraints $E[t_i(X)] = \eta_i$ have smaller entropy: $H(p'(x)) \leq H(p(x))$ with $p(x) = p(x;\theta)$.

Receipe for building MaxEnt Upper Bounds on arbitrary density $q(x)$:

- Choose sufficient statistics $t_i(x)$ so that the differential entropy $H(p(x;\eta))$ of the induced maxent distribution $p(x;\theta)$ is in closed-form (or can be unbounded easily)
- Compute the moment coordinates $\eta_i = E_q[t_i(x)]$, and deduce that $H(q(x)) \leq H(p(x;\eta))$

Absolute Monomial Exponential Family

$$p_l(x;\theta) = \exp\left(\theta|x|^l - F_l(\theta)\right), \quad x \in \mathbb{R} \tag{4}$$

for $\theta < 0$.

Exponential family ($t(x) = |x|^l$) with log-normalizer:

$$F_l(\theta) = \log 2 + \log\Gamma\left(\frac{1}{l}\right) - \log l - \frac{1}{l}\log(-\theta), \tag{5}$$

$\Gamma(u) = \int_0^\infty x^{u-1}\exp(-x)\mathrm{d}x$ generalizes the factorial:
$\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$

Differential entropy of AMEFs

The entropy expressed using the θ-parameter is:

$$\begin{aligned} H_l(\theta) &= \log 2 + \log \Gamma\left(\frac{1}{l}\right) - \log l + \frac{1}{l}(1 - \log(-\theta)), \\ &= a_l - \frac{1}{l}\log(-\theta), \end{aligned} \quad (6)$$

where $a_l = \log 2 + \log \Gamma\left(\frac{1}{l}\right) - \log l + \frac{1}{l}$.

The entropy expressed using the η-parameter is:

$$\begin{aligned} H_l(\eta) &= \log 2 + \log \Gamma\left(\frac{1}{l}\right) - \log l + \frac{1}{l}(1 + \log l + \log \eta), \\ &= b_l + \frac{1}{l}\log \eta, \end{aligned} \quad (7)$$

with $b_l = \log \frac{2\Gamma(\frac{1}{l})(el)^{\frac{1}{l}}}{l}$.

A series of MaxEnt Upper Bounds (MEUBs)

For any continuous RV X, MaxEnt entropy Upper Bound (MEUB) U_l:

$$\boxed{H(X) \leq H_l^{\eta}\left(E_X\left[|X|^l\right]\right)}$$

Are all UBs useful?
That is, can we build a RV X so that $U_{l+1} < U_l$?
(Anwser is yes!)

AMEF MEUBs for Gaussian Mixture Models

Density of a mixture model with k components:

$$m(x) = \sum_{c=1}^{k} w_c p_c(x)$$

Gaussian distribution:

$$p_i(x) = p(x; \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right),$$

$\mu_i = E[X_i] \in \mathbb{R}$: mean parameter
$\sigma_i = \sqrt{E[(X_i - \mu_i)^2]} > 0$: standard deviation

To upper bound $H(X) \leq H_l^\eta\left(E_X\left[|X|^l\right]\right)$, we need to compute the *raw absolute geometric moments* $E_X\left[|X|^l\right]$ for a GMM.

Raw absolute geometric moments of a GMM

Technical part (integration by parts and solving recurrence)

$$A_l(X) = \begin{cases} \sum_{c=1}^{k} w_c \sum_{i=0}^{\lfloor \frac{l}{2} \rfloor} \binom{l}{2i} \mu_c^{l-2i} \sigma_c^{2i} 2^i \frac{\Gamma(\frac{1+2i}{2})}{\sqrt{\pi}} & \\ = \sum_{c=1}^{k} w_c \sum_{i=0}^{\lfloor \frac{l}{2} \rfloor} \binom{l}{2i} \mu_c^{l-2i} \sigma_c^{2i} (2i-1)!! & \text{for even } l, \\ \sum_{c=1}^{k} w_c \sum_{i=0}^{l} \binom{n}{i} \mu_c^{l-i} \sigma_c^{i} \left(I_i\left(-\frac{\mu_c}{\sigma_c}\right) - (-1)^i I_i\left(\frac{\mu_c}{\sigma_c}\right)\right) & \text{for odd } l. \end{cases},$$

where $n!!$ denotes the double factorial: $n!! = \prod_{k=0}^{\lceil \frac{n}{2} \rceil - 1} (n-2k) = \sqrt{\frac{2^{n+1}}{\pi}} \Gamma(\frac{n}{2}+1)$, and:

$$\begin{aligned} I_i(a) &= \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} x^i \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}x, \\ &= \frac{1}{\sqrt{2\pi}} \left(a^{i-1} \exp\left(-\frac{1}{2}a^2\right)\right) + (i-1) I_{i-2}(a), \end{aligned}$$

with the terminal recursion cases:

$$\begin{aligned} I_0(a) &= 1 - \Phi(a) = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{a}{\sqrt{2}}\right)\right) = \frac{1}{2}\mathrm{erfc}\left(\frac{a}{\sqrt{2}}\right), \\ I_1(a) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}a^2\right). \end{aligned}$$

Laplacian MEUB for a GMM ($l = 1$)

AMEF for $l = 1$ is the Laplacian distribution

The differential entropy of a Gaussian mixture model $X \sim \sum_{c=1}^{k} w_c p(x; \mu_c, \sigma_c)$ is upper bounded by:

$$H(X) \leq U_1(X)$$

$$U_{\mathbf{1}}(X) = \log\left(2e\left(\sum_{c=\mathbf{1}}^{k} w_c\left(\mu_c\left(1 - 2\Phi\left(-\frac{\mu_c}{\sigma_c}\right)\right) + \sigma_c\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2}\left(\frac{\mu_c}{\sigma_c}\right)^{\mathbf{2}}\right)\right)\right)\right).$$

Gaussian MEUB for a GMM ($l = 2$)

AMEF for $l = 2$ is the Gaussian distribution

The differential entropy of a GMM $X \sim \sum_{c=1}^{k} w_c p(x; \mu_c, \sigma_c)$ is upper bounded by:

$$H(X) \leq U_2(X) = \frac{1}{2} \log \left(2\pi e \sum_{c=1}^{k} w_c ((\mu_c - \bar{\mu})^2 + \sigma_c^2) \right),$$

with $\bar{\mu} = \sum_{c=1}^{k} w_c \mu_c$.

Vanilla approximation method: Monte-Carlo

Estimate $H(X)$ using *Monte-Carlo (MC) stochastic integration*:

$$\hat{H}_s(X) = -\frac{1}{s}\sum_{i=1}^{s} \log p(x_i), \tag{8}$$

where $\{x_1, \ldots, x_s\}$ is an iid set of variates sampled from $X \sim p(x)$.

MC estimator $\hat{H}_s(X)$ is *consistent*:

$$\lim_{s\to\infty} \hat{H}_s(X) = H(X)$$

(convergence in probability)

However, no deterministic bound, can be above or below true value.

Experiments: Laplacian vs Gaussian MEUBs

$k = 2$ to 10 for $\mu_i, \sigma_i \sim_{\mathrm{iid}} U(0, 1)$, averaged on 1000 trials.

k	Average error	Percentage of times $U_1(X) < U_2(X)$
2	0.5401015778688498	32.7
3	2.7397146972652484	39.2
4	3.4333962273074774	47.9
5	0.9310683623797987	49.9
6	0.5902956910979954	52.1
7	0.7688142345093779	53.2
8	0.2982994538560814	53.8
9	0.1955843679792208	56.8
10	0.1797637053023196	59.9

Important to *recenter the GMMs* so that they have zero expectation (as AMEFs): This does not change the entropy. If not, the 30%+ rates fall significantly to less than 10%.

Are all AMEF MEUBs useful for GMMs?

- For zero-centered GMMs, only Laplacian or Gaussian MEUB is useful,
- For arbitrary GMMs, each bound can be the tightest one ($k = 2$, with GMM mean 0 and two symmetric components with small standard deviation).

Zero-centered GMMs

$U_1(X) < U_2(X)$ iff

$$\log 2e\sqrt{\frac{2}{\pi}}\bar{\sigma_1} \leq \log \sqrt{2\pi e}\overline{\sigma_2}.$$

$$\frac{\overline{\sigma}_1}{\overline{\sigma}_2} \leq \frac{\pi}{2\sqrt{e}} \approx 0.9527$$

$\overline{\sigma}_1$: *arithmetic weighted mean*, $\overline{\sigma}_2 = \sqrt{\sum_{i=1}^k w_i\sigma_i^2}$: *quadratic mean*
weighted quadratic mean dominates *weighted arithmetic mean*: $\frac{\overline{\sigma}_1}{\overline{\sigma}_2} \leq 1$.
$k = 1$: $\sigma > \frac{2\sqrt{e}}{\pi}$ (ie., $\sigma > 1.0496$)

Zero-centered GMMs: $U_{l+2} < U_l$?

Geometric raw (even) moments coincide with the central (even) geometric moments

$$H(X) \leq H_l^\eta(A_l(X)) = b_l + \frac{1}{l} \log z_l + \log \bar{\sigma}_l,$$

$$E_X[X^l] = \underbrace{2^{\frac{l}{2}} \frac{\Gamma(\frac{1+l}{2})}{\sqrt{\pi}}}_{z_l} \left(\sum_{i=1}^k w_i \sigma_i^l \right) = A_l(X).$$

$\bar{\sigma}_l$: *l-th power mean*: $\bar{\sigma}_l = \left(\sum_{i=1}^k w_i \sigma_i^l \right)^{\frac{1}{l}}$

$\frac{\bar{\sigma}_{l+2}}{\bar{\sigma}_l} \geq 1 \Rightarrow \log \frac{\bar{\sigma}_{l+2}}{\bar{\sigma}_l} \geq 0$

$\rightarrow$ not possible (see arXiv).

Arbitrary GMMs: Consider 2-component GMM

$$m(x) = \frac{1}{2}p(x; -\frac{1}{2}, 10^{-5}) + \frac{1}{2}p(x; \frac{1}{2}, 10^{-5})$$

```
H (MC):-9.400517405407735
1       MEUB:0.999999999958284
2       MEUB:0.7257913528258293
3       MEUB:0.5863457882025702
4       MEUB:0.4983017544470345
5       MEUB:0.43651349327316713
6       MEUB:0.390267211711506
7       MEUB:0.35410343073850886
8       MEUB:0.32490700997403515
9       MEUB:0.3007543998901125
10      MEUB:0.2803860698295638
11      MEUB:0.2629389102447494
12      MEUB:0.24779955106708096
13      MEUB:0.23451890956649502
14      MEUB:0.22275989562550735
15      MEUB:0.21226407836562905
16      MEUB:0.2028296978359989
17      MEUB:0.19429672922288133
18      MEUB:0.18653647716356042
19      MEUB:0.17944416377804479
20      MEUB:0.1729335449648154
21      MEUB:0.16693293142890442
22      MEUB:0.16138220185001972
23      MEUB:0.1562305292037145
24      MEUB:0.15143462788690765
25      MEUB:0.14695738668300817
26      MEUB:0.14276679134420478
27      MEUB:0.13883506718452443
28      MEUB:0.1351379906556029 5
29      MEUB:0.13165433203558718
30      MEUB:0.12836540080724268
31      MEUB:0.12525467216646413
```

Contributions and conclusion

- Introduced the class of *Absolute Monomial Exponential Families* (AMEFs) with closed-form log-normalizer,
- Reported closed-form formulæ for the differential entropy of AMEFs,
- Calculated the exact *non-centered absolute geometric moments* for a Gaussian Mixture Model (GMMs),
- Apply MaxEnt Upper Bounds induced by AMEFs to GMMs: All upper bounds are potentially useful for non-centered GMMs
  (But for zero centered-GMMs, only the first two bounds are enough.)
- Recommend $\min(U_1, U_2)$ in applications! (not only U_2)
- Reproducible research with code
  `https://www.lix.polytechnique.fr/~nielsen/MEUB/`

L.D. Brown.
Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory.
IMS Lecture Notes. Institute of Mathematical Statistics, 1986.

Thomas M Cover and Joy A Thomas.
Elements of information theory.
John Wiley & Sons, 2012.

Aapo Hyvarinen, Juha Karhunen, and Erkki Oja.
Independent component analysis.
Adaptive and learning systems for signal processing, communications, and control. John Wiley, New York, Chichester, Weinheim, 2001.

Edwin T Jaynes.
Information theory and statistical mechanics.
Physical review, 106(4):620, 1957.

Edwin T Jaynes.
Information theory and statistical mechanics II.
Physical review, 108(2):171, 1957.

Joseph V Michalowicz, Jonathan M Nichols, and Frank Bucholtz.
Calculation of differential entropy for a mixed Gaussian distribution.
Entropy, 10(3):200–206, 2008.

Joseph Victor Michalowicz, Jonathan M. Nichols, and Frank Bucholtz.
Handbook of Differential Entropy.
Chapman & Hall/CRC, 2013.

Ali Mohammad-Djafari.
A Matlab program to calculate the maximum entropy distributions.
In *Maximum Entropy and Bayesian Methods*, pages 221–233. Springer, 1992.

Frank Nielsen and Richard Nock.
Entropies and cross-entropies of exponential families.
In *2010 IEEE International Conference on Image Processing*, pages 3621–3624. IEEE, 2010.

Sumio Watanabe, Keisuke Yamazaki, and Miki Aoyagi.
Kullback information of normal mixture is not an analytic function.
Technical report of IEICE (in Japanese), pages 41–46, 2004.

Differential entropy of a location-scale family

Density of a *location-scale distribution*: $p(x;\mu,\sigma)=\frac{1}{\sigma}p_0(\frac{x-\mu}{\sigma})$
$\mu\in\mathbb{R}$: *location parameter* and $\sigma>0$: *dispersion parameter*.
Change of variable $y=\frac{x-\mu}{\sigma}$ (with $\mathrm{d}y=\frac{\mathrm{d}y}{\sigma}$) in the integral to get:

$$\begin{aligned} H(X) &= \int_{x=-\infty}^{+\infty} -\frac{1}{\sigma}p_0\left(\frac{x-\mu}{\sigma}\right)\left(\log\frac{1}{\sigma}p_0\left(\frac{x-\mu}{\sigma}\right)\right)\mathrm{d}x,\\ &= \int_{y=-\infty}^{+\infty} -p_0(y)(\log p_0(y)-\log\sigma),\\ &= H(X_0)+\log\sigma. \end{aligned}$$

$\rightarrow$ always independent of the location parameter μ

Non-central even geometric moments of a normal distribution

Even l	$A_l = E\left[\lvert X\rvert^l\right] = E\left[X^l\right] = \sum_{i=0}^{\lfloor \frac{l}{2} \rfloor} \binom{l}{2i}(2i-1)!!\mu^{l-2i}\sigma^{2i}$
2	$\mu^2 + \sigma^2$
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$
8	$\mu^8 + 28\mu^6\sigma^2 + 210\mu^4\sigma^4 + 420\mu^2\sigma^6 + 105\sigma^8$
10	$\mu^{10} + 45\mu^8\sigma^2 + 630\mu^6\sigma^4 + 3150\mu^4\sigma^6 + 4725\mu^2\sigma^8 + 945\sigma^{10}$

Non-central odd geometric moments of a normal distribution

Odd l	$A_l = E\left[|X|^l\right] = C_l(\mu,\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2}) + D_l(\mu,\sigma)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
1	$\sigma\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2}) + \mu\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
3	$(2\sigma^3 + \mu^2\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2}) + (\mu^3 + 3\mu\sigma^2)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
5	$(8\sigma^5 + 9\mu^2\sigma^3 + \mu^4\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2}) + (\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
7	$(48\sigma^7 + 87\mu^2\sigma^5 + 20\mu^4\sigma^3 + \mu^6\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2})+$ $(\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
9	$(384\sigma^9 + 975\mu^2\sigma^7 + 345\mu^4\sigma^5 + 35\mu^6\sigma^3 + \mu^8\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2})+$ $(\mu^9 + 36\mu^7\sigma^2 + 378\mu^5\sigma^4 + 1260\mu^3\sigma^6 + 945\sigma^8)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$

Maxima program

```
assume (theta<0);
F(theta) := log(integrate(exp(theta*abs(x)^5),x,-inf,inf));
integrate(exp(theta*abs(x)^5-F(theta)),x,-inf,inf);
```

Maxima program

```
/* Binomial expansion */
binomialExpansion(i,p,q) := if i = 1 then p+q
else expand((p+q)*binomialExpansion(i-1,p,q)) ;

/* The standard distribution (here, normal) */
p0(y) := exp(-y^2/2)/sqrt(2*pi);

/* Even moment */
absEvenMoment(mu,sigma,l) :=
ratexpand(ratsimp(integrate(factor(expand(binomialExpansion(l,mu,y*sigma)))*p0(y),y,-inf,inf)));


/* Odd moment */
absOddMoment(mu,sigma,l) :=
ratexpand(ratsimp(integrate(factor(expand(binomialExpansion(l,mu,y*sigma)))*p0(y),y,-mu/sigma,inf)
-integrate(factor(expand(binomialExpansion(l,mu,y*sigma)))*p0(y),y,-inf,-mu/sigma)));


/* General : Maxima does not give a closed-form formula
because of the absolute value */
absMoment(mu,sigma,l) :=
ratexpand(ratsimp(integrate(abs(factor(expand(binomialExpansion(l,mu,y*sigma))))*p0(y),y,-inf,inf)));


assume(sigma>0);
assume(mu>0);  /* maxima needs to branch condition */
absEvenMoment(mu,sigma,8);
absOddMoment(mu,sigma,7);
```

(%o5) absMoment (μ,σ,l) :=

ratexpand(ratsimp($\int_{-\infty}^{\infty}$ |factor (expand (binomialExpansion (l,μ,y σ)))| p0 (y) dy))

(%o6) [σ>0]

(%o7) [μ>0]

(%o8) $$\frac{105\sqrt{\pi}\,\sigma^{8}}{\sqrt{\pi}}+\frac{420\sqrt{\pi}\,\mu^{2}\sigma^{6}}{\sqrt{\pi}}+\frac{210\sqrt{\pi}\,\mu^{4}\sigma^{4}}{\sqrt{\pi}}+\frac{28\sqrt{\pi}\,\mu^{6}\sigma^{2}}{\sqrt{\pi}}+\frac{\sqrt{\pi}\,\mu^{8}}{\sqrt{\pi}}$$

(%o9) $$\frac{3\,2^{9/2}\sigma^{7}\,\%e^{-\frac{\mu^{2}}{2\sigma^{2}}}}{\sqrt{\pi}}+\frac{87\sqrt{2}\,\mu^{2}\sigma^{5}\,\%e^{-\frac{\mu^{2}}{2\sigma^{2}}}}{\sqrt{\pi}}+\frac{5\,2^{5/2}\mu^{4}\sigma^{3}\,\%e^{-\frac{\mu^{2}}{2\sigma^{2}}}}{\sqrt{\pi}}+\frac{\sqrt{2}\,\mu^{6}\sigma\,\%e^{-\frac{\mu^{2}}{2\sigma^{2}}}}{\sqrt{\pi}}+\frac{105\sqrt{\pi}\,\mu\,\mathrm{erf}\left(\frac{\mu}{\sqrt{2}\,\sigma}\right)\sigma^{6}}{\sqrt{\pi}}$$
$$+\frac{105\sqrt{\pi}\,\mu^{3}\,\mathrm{erf}\left(\frac{\mu}{\sqrt{2}\,\sigma}\right)\sigma^{4}}{\sqrt{\pi}}+\frac{21\sqrt{\pi}\,\mu^{5}\,\mathrm{erf}\left(\frac{\mu}{\sqrt{2}\,\sigma}\right)\sigma^{2}}{\sqrt{\pi}}+\frac{\sqrt{\pi}\,\mu^{7}\,\mathrm{erf}\left(\frac{\mu}{\sqrt{2}\,\sigma}\right)}{\sqrt{\pi}}$$