On determination of function extrema with MaxEnt formulation

Ravi Sankar Saripalli

May 29, 2019

Description

This is an attempt to find extrema of a function with MaxEnt formulation. The usual way to find extrema of a function $f(x)$ is to find x where function derivative $f'(x)$ vanishes. An alternate approach is to use probability distribution function $p(x)$ as an independant parameteric function and seek to maximize the expected value of $f(x)$ based on the $p(x)$ with the condition that entropy (as defined by Shanon) of the probability distribution $p(x)$ is maximized, thus ensuring that the distribution function derived is least biased.

The MaxEnt Lagrangian

$$\boxed{\mathcal{L}(p,\lambda) = \int f(x)p(x)dx + T\int p(x)ln(p(x))dx + \lambda\left\{\left(\int p(x)dx\right) - 1\right\}} \tag{1}$$

The first term in the Lagrangian is the expected value of the function corresponding to the PDF $p(x)$, the second term corresponds to the entropy of the PDF scaled by an arbitray constant T and the last term corresponds to the equality constraint that integral of PDF is one. The multiplier λ is the Lagrangian parameter.
Noting that the Lagrangian is function of $p(x)$ and λ, the extrema of the Lagrangian is obtained by requiring the functional derivative with respect $p(x)$ and the Lagrangian parameter λ are zero.
Although this enables us to determine $p(x)$ and λ corresponding to the extrema, we need additional criterion to determine if the extrema is either minimum or maximum. Sign of the second derivative in functional space similar to simple variables can be used to ascertain if the stationary point corresponds to minima or maxima.

The Euler-Lagrange equation

Consider the following functional (function of functions)

$$\boxed{F\left(f', f, x\right) = \int \phi\left(f'(x), f(x), x\right)dx} \tag{2}$$

Requiring that functional derivative of F with respect to f is zero at extrema (note f is a function not a variable, hence the name functional derivative) following Eauler-Lagrange equation can be derived.

$$\boxed{\frac{\partial\phi}{\partial y} - \frac{d}{dx}\left(\frac{\partial\phi}{\partial y'}\right) = 0} \tag{3}$$

Derivation of the above equation that is easy to follow is here

Finding Stationary Point of Lagrangian in Eq 1.

Using the above Euler-Lagrange equation, the stationary conditions for the Lagrangian in equation 1, can be derived from functional derivative with respective p and the derivative with respective λ as follows.

$$\boxed{f(x) + T\,\{1 + ln(p(x))\} + \lambda = 0} \tag{4}$$

Setting derivative with respect to λ to zero yields the contraint that PDF integral should be 1.

$$\boxed{\int p(x)\,dx - 1 = 0} \tag{5}$$

Rearranging eqn.4

$$\boxed{p(x) = e^{-(1+\lambda/T)}\,e^{-f(x)/T}} \tag{6}$$

Combining eqns. 5 and 6 the value of lambda can be expressed as follows

$$\boxed{\lambda = T\left\{ln \int e^{-f(x)/T} dx - 1\right\}} \tag{7}$$

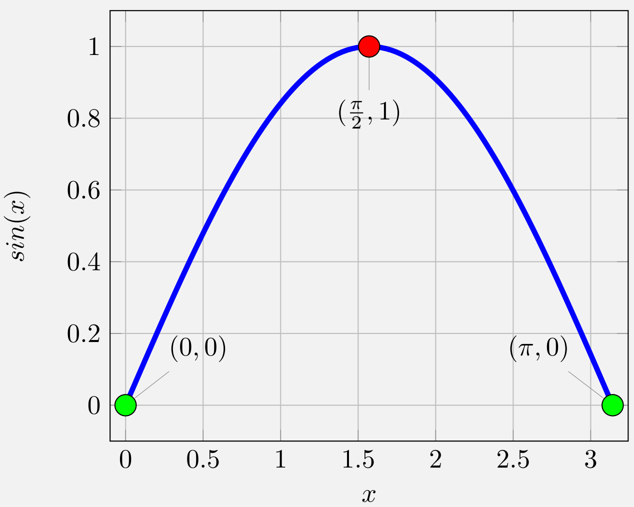
The question to ask is how do we now proceed to get min or max of $f(x)$ given that we have it in analytical form (eg. sin(x)).. What will be the p(x) as T approaches zero. That is the intent of the annealing process as we converge on p(x).

Explorations with sinx

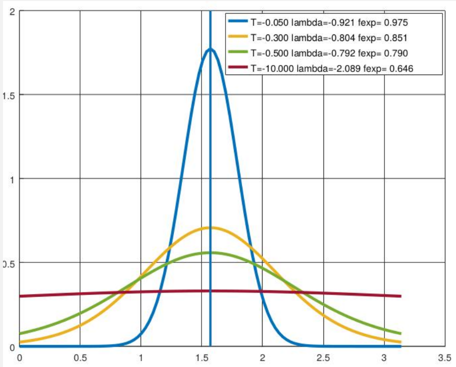
Assuming that we are interested in finding extrema of sin(x) in the interval $[0, \pi]$, one can proceed as follows.

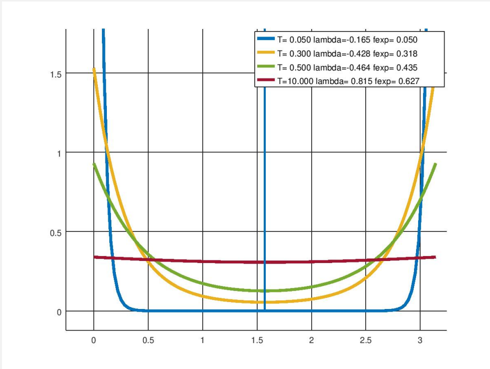
1. For a given T, calculate λ by evaluating the integral numerically
2. With $Tand\lambda$ known, the PDF $p(x)$ now is well defined (eq.6). The expected value of $f(x)$ can be determined by numerical integration of $\int p(x)f(x)dx$

$sin(x)$



Change in pdf with T when $T \leq 0$ and $T \geq 0$

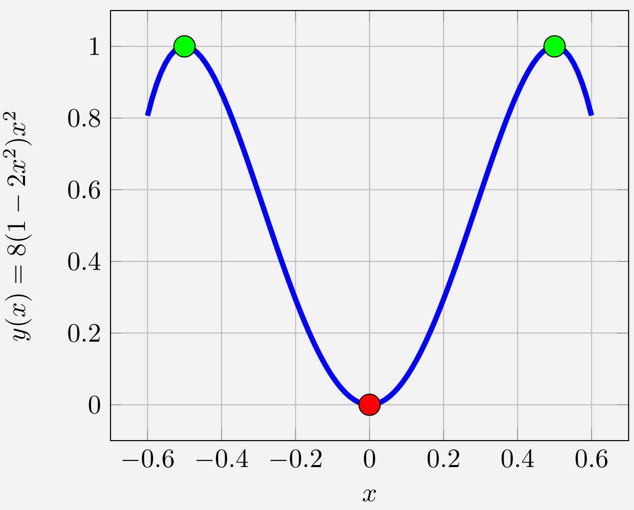




From the above figures, it is clear that for $T \geq 0$ as its magnitude approaches zero, the pdf spread decreases and tends to peak around the true maxima. On the otherhand, when $T \geq 0$, the pdf peaks appear near the end points where the function has minimal value in the interval of interest.

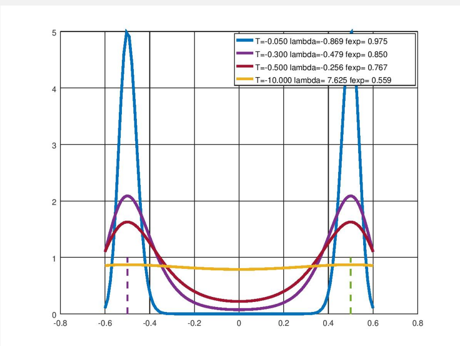
Let us now explore a different function that has two peaks, and trough as shown in the following figure.

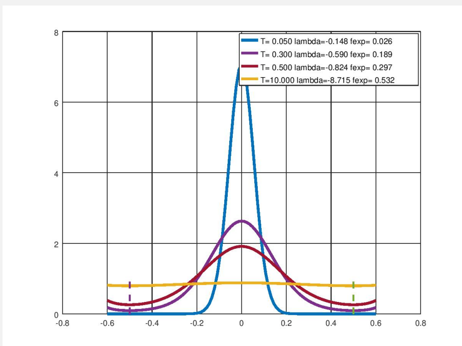
$y(x) = 8(1 - 2x^2)x^2 - 0.6 < x < 0.6$



Using the above function, the effect of T on PDF corresponding to the stationary point is determined as described earlier.

Effect of T on PDF at stationary condition





Once again, when T value is approaching zero from positive side, the PDF corresponding to extrema condition peaks around the true minima at $x = 0$ with corresponding expected value of function approaching 0. When T approaches zero from the negative side, the PDF peaks around $x = -0.5, and 0.5$ with expected value of function approaching 1.

It therefore appears that when T approaches zero from negative side PDF determined from stationary condition corresponds to maxima of the function, while it approaches zero from positive side, the PDF peaks around the minima of the function.

Functional Derivatives

In this section the derivation of functional derivatives is provided. While there is no shortage of articles on this topic, thesis work by Aaron Neson provides lucid explanation.
Let $\eta(x)$ be an arbitrary function which vanishes to zero at the boundaries, and ϵ is an arbitrarily small constant. For the functional F defined in equaiton 2, we denote its gradient with respect to $f(x)$ with ∇F_f, and $\eta(x)$ is the variation of function $f(x)$, then

$$\boxed{\begin{aligned} &\langle \nabla F_f\,, \eta \rangle = \frac{d}{d\epsilon} F\left(f + \epsilon\eta\right)\Big|_{\epsilon=0} \\ &\text{where} \\ &\langle \nabla F_f, \eta \rangle = \int \nabla F_f \eta(x) dx \end{aligned}} \tag{8}$$

$$\begin{aligned} \langle \nabla F_f\,, \eta \rangle &= \frac{d}{d\epsilon} \int \phi\left(f + \epsilon\eta, f' + \epsilon\eta', x\right) dx\Big|_{\epsilon=0} \\ &= \int \left\{ \eta \frac{\partial \phi\left(f + \epsilon\eta, f' + \epsilon\eta', x\right)}{\partial\left(f + \epsilon\eta\right)} + \eta' \frac{\partial \phi\left(f + \epsilon\eta, f' + \epsilon\eta', x\right)}{\partial\left(f' + \epsilon\eta'\right)} \right\} dx\Big|_{\epsilon=0} \\ &= \int \left\{ \eta \frac{\partial \phi\left(f, f', x\right)}{\partial f} + \eta' \frac{\partial \phi\left(f, f', x\right)}{\partial f'} \right\} dx \\ &= \int \eta \frac{\partial \phi\left(f, f', x\right)}{\partial f} dx + \cancelto{0}{\left[\eta \frac{\partial \phi\left(f, f', x\right)}{\partial f} \right]_a^b} - \int \eta \frac{d}{dx} \left\{ \frac{\partial \phi\left(f, f', x\right)}{\partial f'} \right\} dx \\ &= \int \eta \left[\frac{\partial \phi\left(f, f', x\right)}{\partial f} - \frac{d}{dx} \left\{ \frac{\partial \phi\left(f, f', x\right)}{\partial f'} \right\} \right] dx \end{aligned}$$

Since $\eta(x)$ is arbitray function

$$\boxed{\nabla F_f = \frac{\partial \phi\left(f, f', x\right)}{\partial f} - \frac{d}{dx} \left\{ \frac{\partial \phi\left(f, f', x\right)}{\partial f'} \right\}} \tag{9}$$

Similarly the second variation of f can be defined as follows

$$\boxed{\begin{aligned} &\left\langle \nabla^2 F_f\,, \eta \right\rangle = \frac{d^2}{d\epsilon^2} F\left(f + \epsilon\eta\right)\Big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon} \left\langle \nabla F_f\,, \eta \right\rangle\Big|_{\epsilon=0} \end{aligned}} \tag{10}$$