Individualized Extended J-QPD Regression

M. Feindt, F. Wick

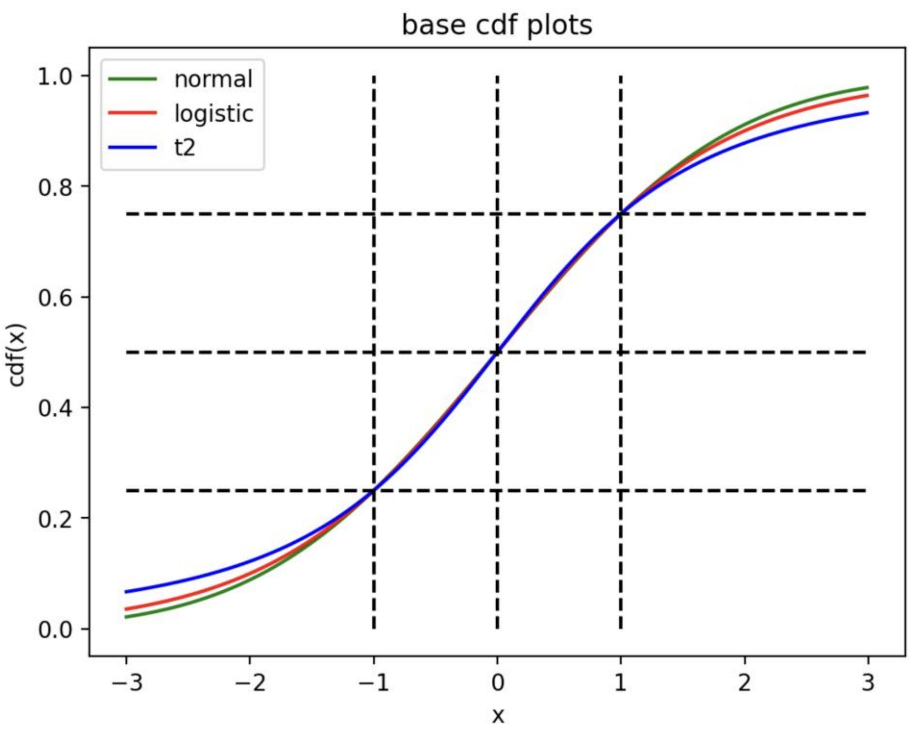
December 2023

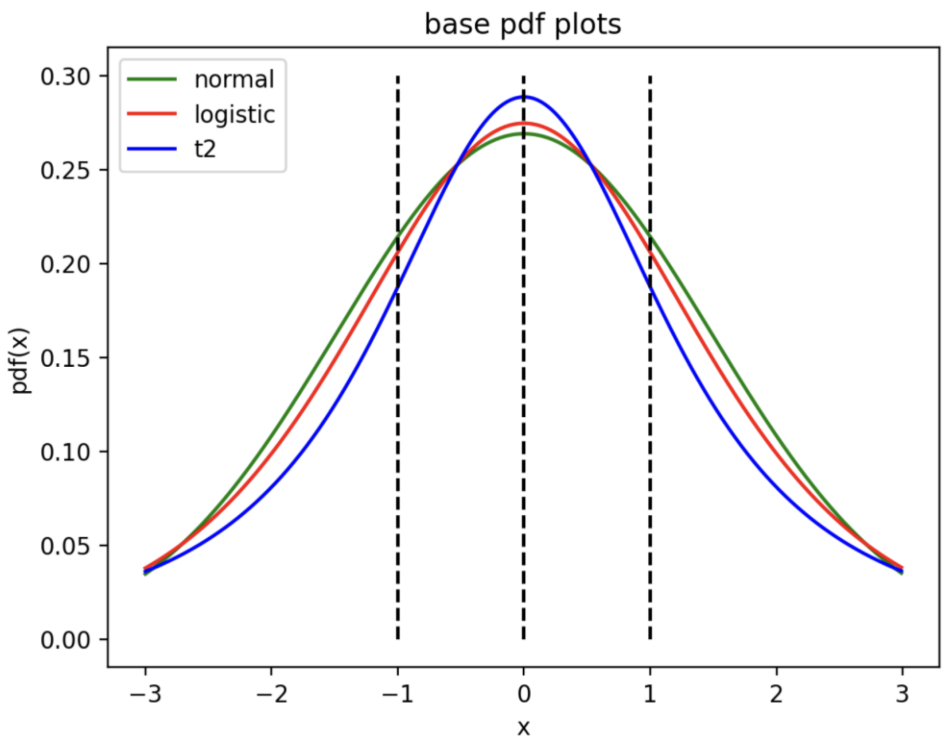
Base distributions

We define the scale of base distributions such that always

$$Q(\alpha) = -1\,, Q(0.5) = 0, Q(1-\alpha) = 1$$

independent of the chosen distribution itself. Then they look as follows (here α=0.25). The tail behaviour will be determined by the choice of the base distribution. The JQPD parameter calculation will not depend on the choice (since it only depends on the 3 symmetric quantiles.)



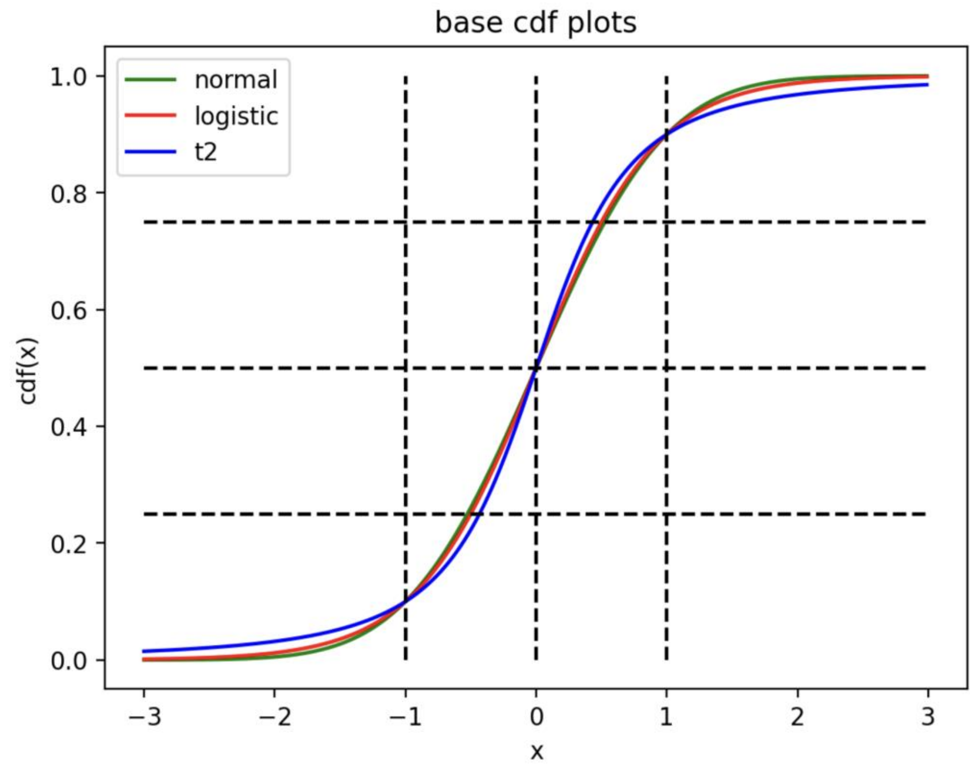


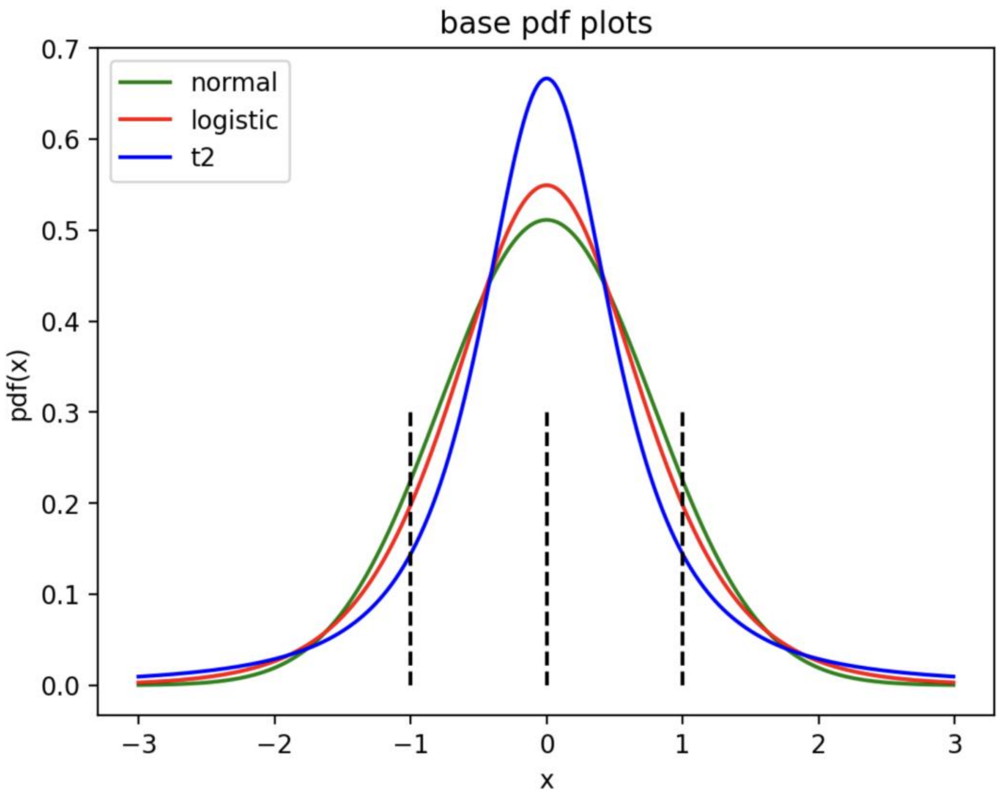
Base distributions II

We define the scale of base distributions such that always

$$Q(\alpha) = -1\,, Q(0.5) = 0, Q(1-\alpha) = 1$$

independent of the chosen distribution itself. Then they look as follows (here for α=0.10). The tail behaviour will be determined by the choice of the base distribution. The JQPD parameter calculation will not depend on the choice of the distribution but on the choice of α.



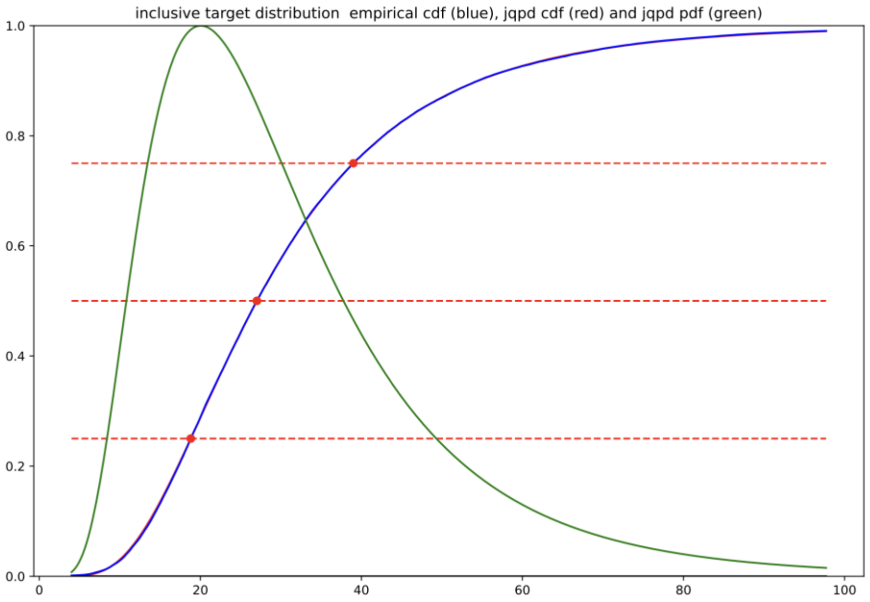


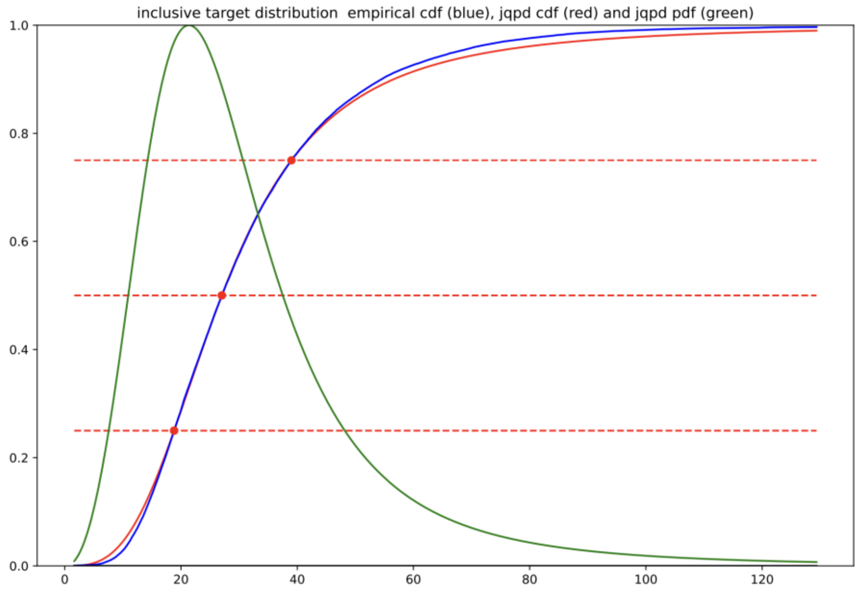
Base distributions III

Idea:

We can determine a good choice of the base distribution by calculating different JQPDs for the inclusive target distribution and make sure a distribution is chosen that well fits the empiric cdf.

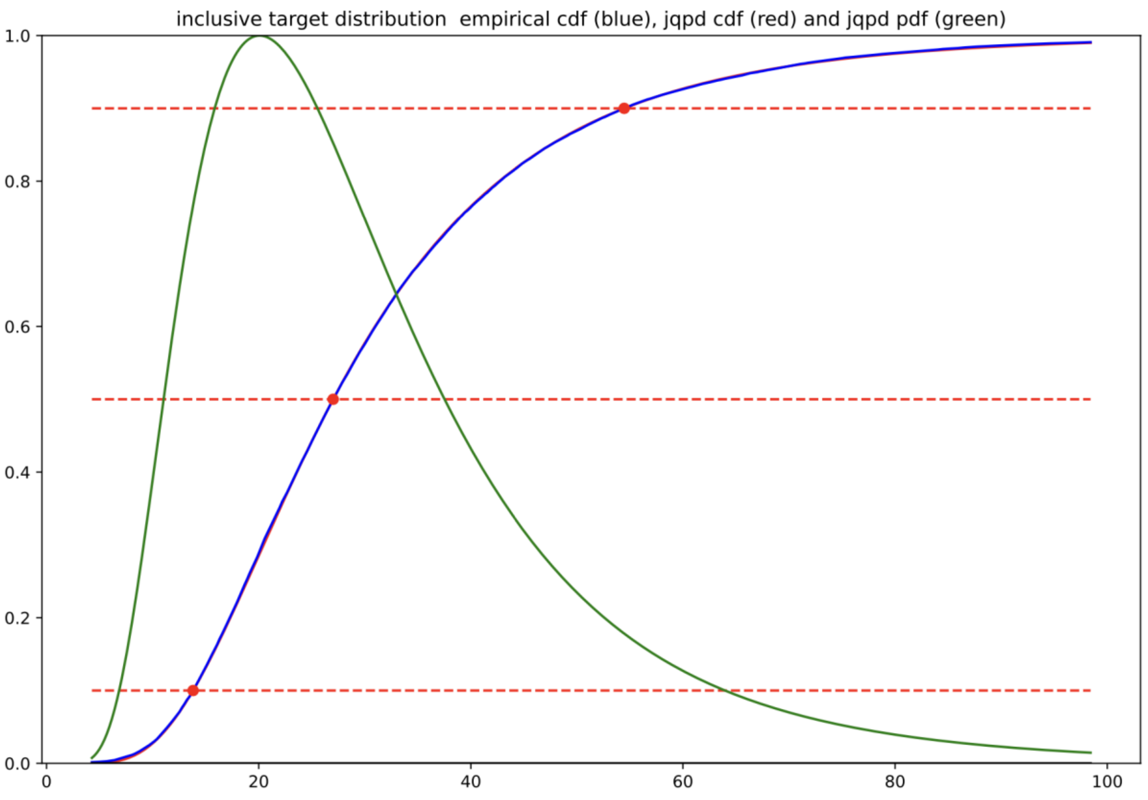
Example: Perfect fit here with normal base distribution (left).
Too strong tails predicted with logistic distribution (right)

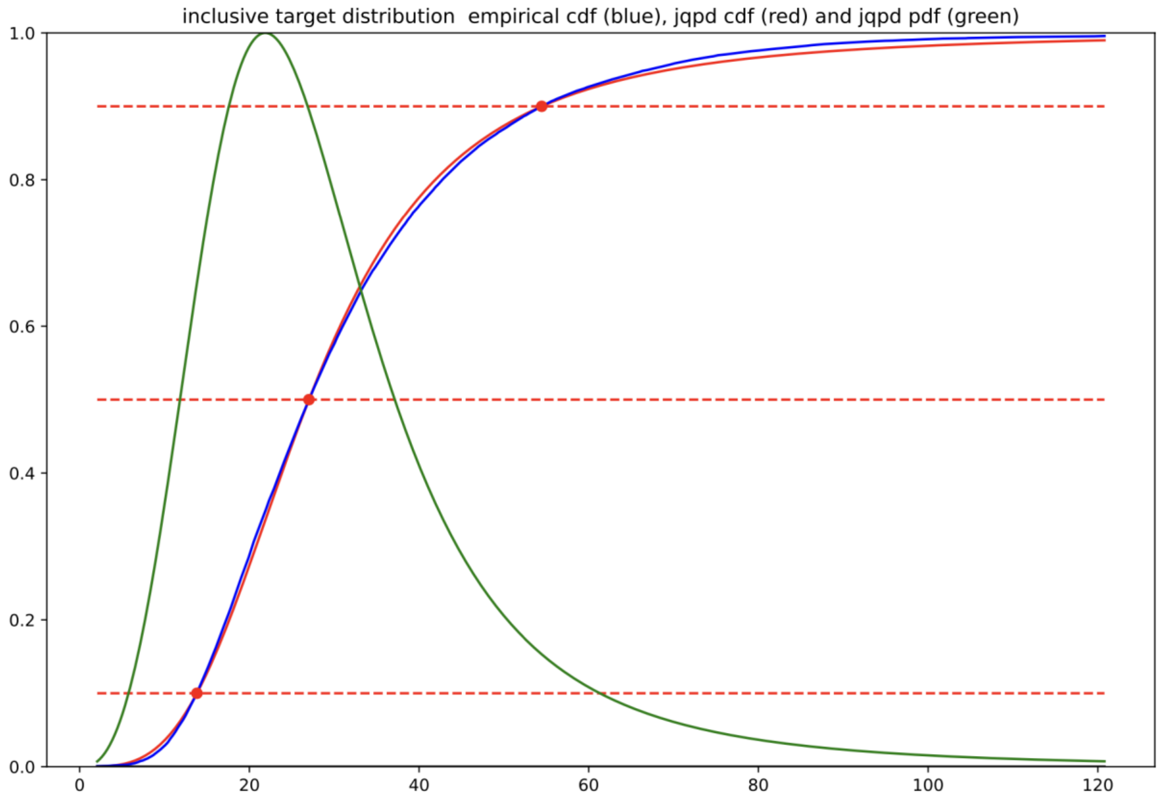




Base distributions IV

How does it look when we use the 90% quantile?
Same example: Perfect fit again with normal base distribution (left), also with $\alpha = 0.1$ quantile.
Still too strong tails predicted with logistic distribution (right).
Logistic is the worse fit in this case. We can fit a continuous parameter interpolating between different distributions.





Bijective transformations for u

We define transformations that map the physical variables x with constraints (e.g., positive or in interval) to internal variables z defined in whole R with no constraints.

For unconstrained variables T may be:

$z = T(x) = x$
$z = T(x) = \sinh(x)$

and the corresponding inverse transformations:

$\mathrm{x} = T^{-1}(z) = z$
x= $T(z) = \operatorname{arcsinh}(z)$

We first will take the identity.

Bijective transformations for b

We define transformations T that map the physical variables x with the constraint to reside in an interval [lb,ub] to internal variables z defined in whole R with no constraints.

For bounded variables T may consist of 2 steps:

$$\mathrm{p}(x) = \frac{x-lb}{ub-lb}$$ (mapping [lb,ub] to [0,1])

$$z = T(x) = \log\left(\frac{p(x)}{1-p(x)}\right) = log\,\frac{x-lb}{ub-x}$$ (logistic ppf (logit) or a ppf of any other distribution)

The inverse transformation reads:

$$\mathrm{p}(z) = \frac{1}{1+\exp(-z)}$$ (logistic cdf function or the cdf of the other distribution chosen)

$$x = T^{-1}\,(z) = lb + (ub - lb) * p(z)$$ (mapping from [0,1] to [lb,ub])

Bijective transformations for s

We define transformations T that map the physical variables x with the constraint tobe larger than a lower bound Lb (an often occuring value being 0) to internal variables z defined in whole R with no constraints.

For semibounded variables T may consist of 2 steps:

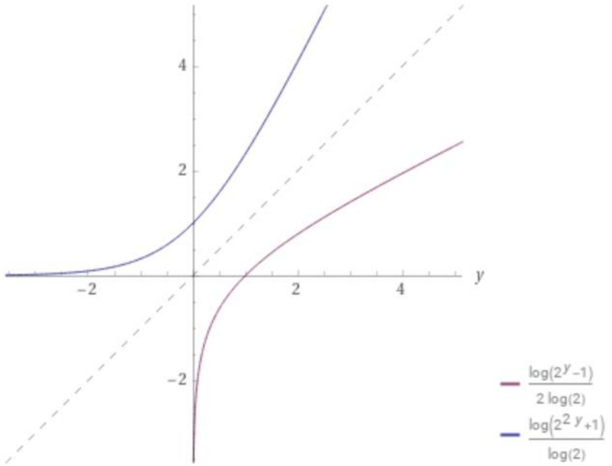
$pos(x) = x - lb$ (mapping $[lb, \infty)$ to $[0, \infty)$)
$z = T(x) = \log\big(pos(x)\big)$ (logarithm)

The inverse transformation reads:

$pos(z) = \exp(z)$ (logistic cdf function or the cdf of the other distribution chosen)
$x = T^{-1}\,(z) = lb + pos(z)$ (mapping from [0,1] to [lb,ub])

Alternative: softplus instead of logarithm:
$z = T(x) = $ log(2(exp(y)−1))/(2∗log(2))

Leads to additive instead of multiplicative behaviour away from threshold



Extended J-QPD transform

An extended J_QPD transform class is defined by

- *α, defining the low quantile of the triplet defining the J-QPD e.g. 0.25 (the high quantile is $1-\alpha$)*
- *Base distribution (symmetric with median 0 ,i.e. $\Phi^{-1}(0.5)$=0), like normal or logistic or an interpolation of these, with cdf p=$\Phi(x)$ and its inverse, the base quantile function $Q=\Phi^{-1}(p)$. The scale of a base distribution is defined by the condition $\Phi^{-1}(1-\alpha)$=1, (or equivalently, due to the symmetry , $\Phi^{-1}(\alpha)$=-1), see next page.*
- *Possible constraints*
    - *unbounded: bounds= u, b_l = None, b_h =None,*
        - *transformation T(x)=x*
    - *semi-bounded: bounds=s, b_l = lower bound, often 0, b_u =None*
        - *transformation: logarithm, T(x)=log(x)*
    - *interval bounded: bounds =b, b_l = lower bound, often 0, b_u =upper bound, often 1*
        - *transformation: logit, T(x)=log(x/(1-p))*

- *Possibly other transformations in internal unconstrained space may be useful, this will be explored later.*

In a training one specific J-QPD transform class will be used, e.g. one which decribes well the inclusive distribution.
A specific member is then given by the three quantile values in physical space. These can be transformed.

Base distributions for extended J-QPD

We define the scale of base distributions such that always

$$Q(\alpha) = -1\,,Q(0.5) = 0, Q(1-\alpha) = 1$$

independent of the chosen distribution itself (see chapter base distributions)
Observe that the definition depends on α.

Here are normal and logistic base distributions as examples:

Normal: (using scipy.stats)

```
def basenormppf(alpha,p):
    wid=norm.ppf(1-alpha)
    return norm.ppf(p)/wid

def basenormcdf(alpha,x):
    wid=norm.ppf(1-alpha)
    return norm.cdf(x*wid)

def basenormpdf(alpha,x):
    wid=norm.ppf(1-alpha)
    return norm.pdf(x*wid)*wid
```

Logistic: (explicit coding)

```
def baselogisticppf(alpha,p):
    wid=np.log((1-alpha)/alpha)
    return np.log(p/(1-p))/wid

def baselogisticcdf(alpha,x):
    wid=np.log((1-alpha)/alpha)
    return 1./(1+np.exp(-x*wid))

def baselogisticpdf(alpha,x):
    wid=np.log((1-alpha)/alpha)
    return wid/(np.exp(x*wid/2)+np.exp(-x*wid/2))**2
```

Extended J-QPD quantile and cdf functions

The quantile function x=ppf(p) in physical (constrained) space x can be constructed from its probability p from this chain:

1. Calc base quantile q_b from base distribution: $q_b = \Phi_b^{-1}(p)$
2. Calc Johnson S_U transform: $z = \xi + \lambda \sinh(\frac{q_b - \gamma}{\delta})$
3. Transform into physical (constrained) space using inverse T: $x = T^{-1}(z)$

The cdf function p=cdf(x) from physical (constrained) space x can be constructed from the inverse: mapping:

1. Transform from physical (constrained) space using: $z = \mathrm{T}(x)$
2. Calc inverse Johnson S_U transform: $q_b = \gamma + \delta \sinh^{-1}(\frac{z-\xi}{\lambda})$
3. Calc probability p from base quantile: $p = \Phi_b(q_b)$

The pdf function p=pdf(x) in physical (constrained) space x can be calculated by differentiating the cdf, either analytically or numerically)

Solution from 3 symmetric quantiles

The 3 symmetric quantiles in base quantile space are by definition:

$q_{b,low} = \Phi_b^{-1}(\alpha) = -1$, $q_{b,med} = \Phi_b^{-1}(0.5) = 0$, $q_{b,high} = \Phi_b^{-1}(1-\alpha) = 1$

The corresponding values in unconstrained z-space are

$$z_l, z_m, z_h$$

Then a solution for the internal Johnson-parameters can be found via:

$$\frac{z_l - \xi}{\lambda} = \sinh\left(\frac{-\gamma - 1}{\delta}\right) \qquad \frac{z_m - \xi}{\lambda} = \sinh\left(\frac{-\gamma}{\delta}\right) \qquad \frac{z_h - \xi}{\lambda} = \sinh\left(\frac{-\gamma + 1}{\delta}\right)$$

Fix γ *(is possible since 4 free parameters are only constrained by 3 conditions) to be* $q_{b,high} = 1$*.*
We will see that this works if $z_h - 2z_m + z_l < 0$ *(the case that in the literature is called n=-1).*

$$\frac{z_l - \xi}{\lambda} = \sinh\left(\frac{-2}{\delta}\right) \qquad \frac{z_m - \xi}{\lambda} = \sinh\left(\frac{-1}{\delta}\right) \qquad \frac{z_h - \xi}{\lambda} = \sinh\left(\frac{0}{\delta}\right)$$

The 3. equation implies $\xi = z_h$*.*

Solution from 3 symmetric quantiles II

Thus the two remaining equations read (taking sinh asymmetry into account). The third just is 0=0

$$\frac{z_h - z_l}{\lambda} = \sinh\left(\frac{2}{\delta}\right) \qquad \frac{z_h - z_m}{\lambda} = \sinh\left(\frac{1}{\delta}\right)$$

Note that the 2 sinh arguments differ by exactly a factor of 2, and a sinh addition theorem will help us in the following.

To get rid of λ*, we take the calculable ratio (and call it A) to achieve*

$\frac{z_h - z_m}{z_h - z_l} = A = \frac{\sinh\left(\frac{1}{\delta}\right)}{\sinh\left(\frac{2}{\delta}\right)} = \frac{\sinh\left(\frac{1}{\delta}\right)}{2\sinh\left(\frac{1}{\delta}\right)\cosh\left(\frac{1}{\delta}\right)} = \frac{1}{2\,cosh\left(\frac{1}{\delta}\right)}$.

We can solve for δ: $\delta = \frac{1}{\operatorname{arccosh}\left({}^{1}/_{2A}\right)}$

It has a solution when the arccosh argument is larger than 1, i.e. A<1/2

This is the case when $z_h - z_m < (z_h - z_l)/2$ *which is equivalent to* $z_h - 2z_m + z_l < 0$ $(case\ n < 0)$

Then

$\lambda = (z_h - z_l)/\sinh\left(\frac{2}{\delta}\right)$.

Solution from 3 symmetric quantiles III

Now consider the n=+1 case $z_h - 2z_m + z_l > 0$.

In this case fix γ *to be* $q_{b,low} = -1$.

$$\frac{z_l - \xi}{\lambda} = \sinh\left(\frac{0}{\delta}\right) \qquad \frac{z_m - \xi}{\lambda} = \sinh\left(\frac{1}{\delta}\right) \qquad \frac{z_h - \xi}{\lambda} = \sinh\left(\frac{2}{\delta}\right)$$

The 1. equation implies $\xi = z_l$.

Thus the two remaining equations read (taking sinh asymmetry into account). The first just is 0=0)

$$\frac{z_h - z_l}{\lambda} = \sinh\left(\frac{2}{\delta}\right) \qquad \frac{z_m - z_l}{\lambda} = \sinh\left(\frac{1}{\delta}\right)$$

Note that the 2 sinh arguments again differ by exactly a factor of 2, and we apply the same sinh addition theorem.

Solution from 3 symmetric quantiles IV

To get rid of λ*, we take the calculable ratio (and call it A) to achieve*

$$\frac{z_m - z_l}{z_h - z_l} = A = \frac{\sinh\left(\frac{1}{\delta}\right)}{\sinh\left(\frac{2}{\delta}\right)} = \frac{\sinh\left(\frac{1}{\delta}\right)}{2\sinh\left(\frac{1}{\delta}\right)\cosh\left(\frac{1}{\delta}\right)} = \frac{1}{2\,cosh\left(\frac{1}{\delta}\right)}.$$

We can solve for δ: $\delta = \frac{1}{\operatorname{arccosh}\left({}^{1}/_{2A}\right)}$

It has a solution when the arccosh argument is larger than 1, i.e. A<1/2

This is the case when $z_h - z_m < (z_h - z_l)/2$ *which is equivalent to* $z_h - 2z_m + z_l < 0$.

Then

$\lambda = (z_h - z_l)/\sinh\left(\frac{2}{\delta}\right)$.

In the case n=0, i.e. the symmetric case $z_h - 2z_m + z_l = 0$*, there is no need for an additional non-linearity and only a linear shift / scale should be used:*

ppf: $z = \xi + \lambda\, q_b$

cdf: $q_b = (z - \xi)/\lambda$

with $\xi = z_m$ *and* $\lambda = z_h - z_m$

Solution algorithm summary

$\gamma = -sign\ (\ z_h - 2z_m + z_l\ )$
if $\gamma > 0$:

$\xi = z_l$

$A = \frac{z_m - z_l}{z_h - z_l}$

else if $\gamma < 0$:

$\xi = z_h$

$A = \frac{z_h - z_m}{z_h - z_l}$

$\delta = \frac{1}{\operatorname{arccosh}(1/_{2A})}$

$\lambda = (z_h - z_l) / \sinh\left(\frac{2}{\delta}\right)$.

For n=0 (symmetric case)
replace the Johnson transform by an affine linear transform:

ppf: $z \ = \xi + \lambda\, q_b$

cdf: $q_b = (z - \ \xi)/\lambda$

with
$\xi = z_m$
$\lambda = \ z_h - z_m$