The Fitted Gamma Function

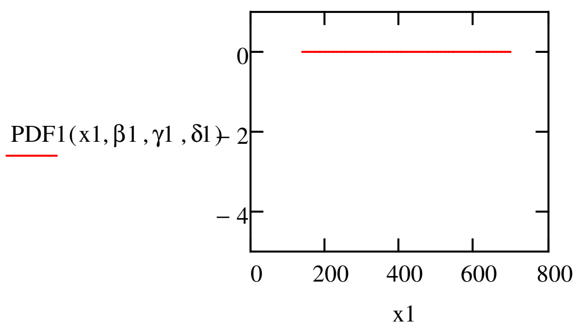
This worksheet demonstrates the shape and behavior of the Gamma Distribution and fits data to a three parameter version of the function. Here is the PDF function that is to be fitted:

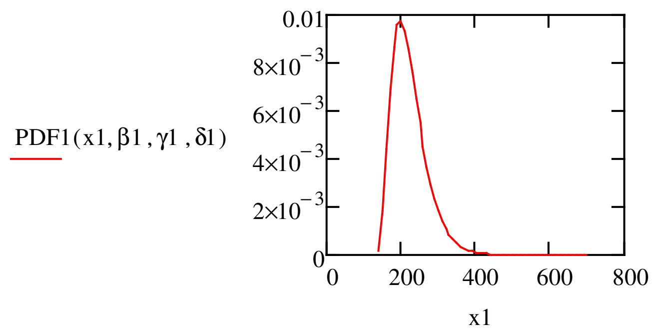
$\beta1 := 25 \qquad \gamma1 := 3.5 \qquad \delta1 := 135 \qquad x1 := 0, 10 .. 700$ Typical values

$$PDF1(x1, \beta1, \gamma1, \delta1) := \frac{\left(\frac{x1 - \delta1}{\beta1}\right)^{\gamma1 - 1} \cdot e^{-\frac{x1 - \delta1}{\beta1}}}{\beta1 \cdot \Gamma(\gamma1)}$$

Offset Gamma Function

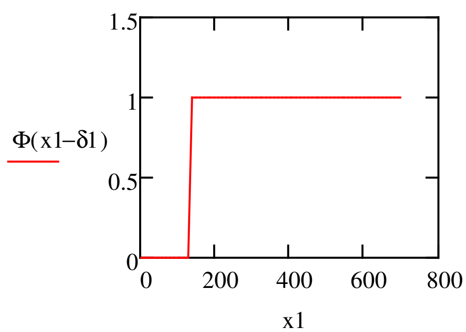
beta is a scale parameter; gamma is a shape parameter and delta is the offset from zero where the function "starts". Here is a typical example of this curve:





The PDF given above behaves incorrectly for x less than 135. The PDF function value needs to be set to zero for x less than delta (135). This construct is one way to do it:

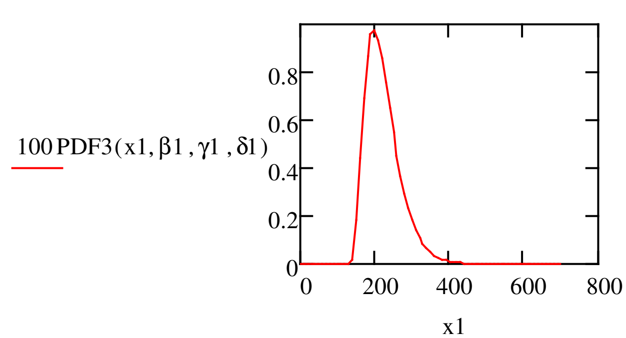
$$PDF2(x1, \beta2, \gamma2, \delta2) := \text{if}(x1 > \delta2, PDF1(x1, \beta1, \gamma1, \delta1), 0)$$



We can also use the "step function" to the left to force the PDF curve to equal zero for x less than delta:

$$PDF3(x1, \beta1, \gamma1, \delta1) := PDF1(x1, \beta1, \gamma1, \delta1) \cdot \Phi(x1 - \delta1)$$

PDF3 is more tractable for subsequent use than is PDF2



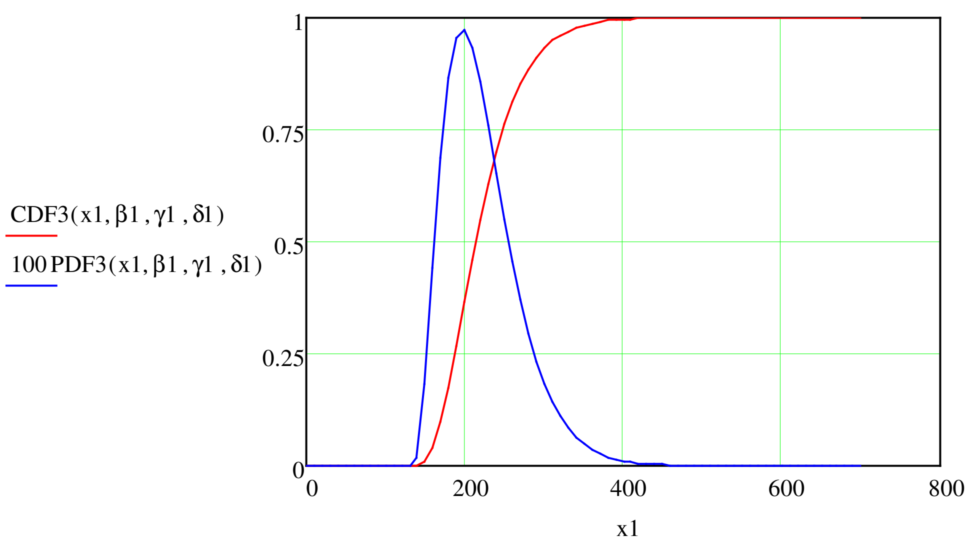
The next hurdle is that though the PDF to be fit is the PDF2 given above ... what we need to do is fit the integral of PDF2 ... the cumulative distribution function (CFD). The CFD is the data presented in available journal papers and The Book of Myrtle.

The CDF is the definite integral of PDF2 from 0 to x

$$k := 0, 1 .. 59$$

$$CDF3(x1, \beta1, \gamma1, \delta1) := \int_0^{x1} PDF2(z, \beta1, \gamma1, \delta1)\, dz$$

The function to be fitted is the cumulative distribution function



The above work up uses typical values for the parameters that define the PDF and the CDF. Now we do it all over again ... this time with real data.

Here is the data to be fit. It is from The Book of Myrtle Fig 17

$$data := \begin{pmatrix} 0 & 0 \\ 30 & 0 \\ 60 & 0 \\ 90 & 0 \\ 120 & 0 \\ 150 & 0 \\ 180 & 0 \\ 210 & 0 \\ 240 & 0 \\ 270 & 0 \\ 300 & 0 \\ 330 & 0.11 \\ 360 & 0.1 \\ 390 & 0.33 \\ 420 & 0.5 \\ 450 & 0.57 \\ 510 & 1 \\ 540 & 1 \\ 570 & 1 \\ 600 & 1 \\ 630 & 1 \\ 700 & 1 \end{pmatrix}$$

We split the data vector into two vectors ... one for x and one for y:

$x := data^{\langle 0 \rangle}$ $\quad$ $y := data^{\langle 1 \rangle}$

$length(x) = 22$ $\quad$ $Range := length(x)$

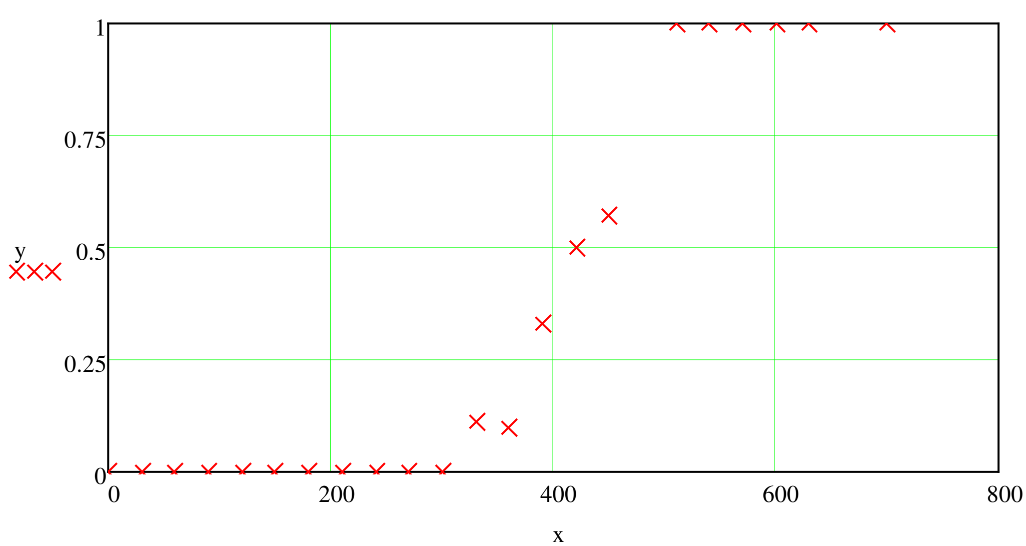
$n1 := length(x) - 1$ $\quad$ $n1 = 21$

$Range = 22$

data =

	0	1
0	0	0
1	30	0
2	60	0
3	90	0
4	120	0
5	150	0
6	180	0
7	210	0
8	240	0
9	270	0
10	300	0
11	330	0.11
12	360	0.1
13	390	0.33
14	420	0.5
15	450	...

Here is a plot of this data:



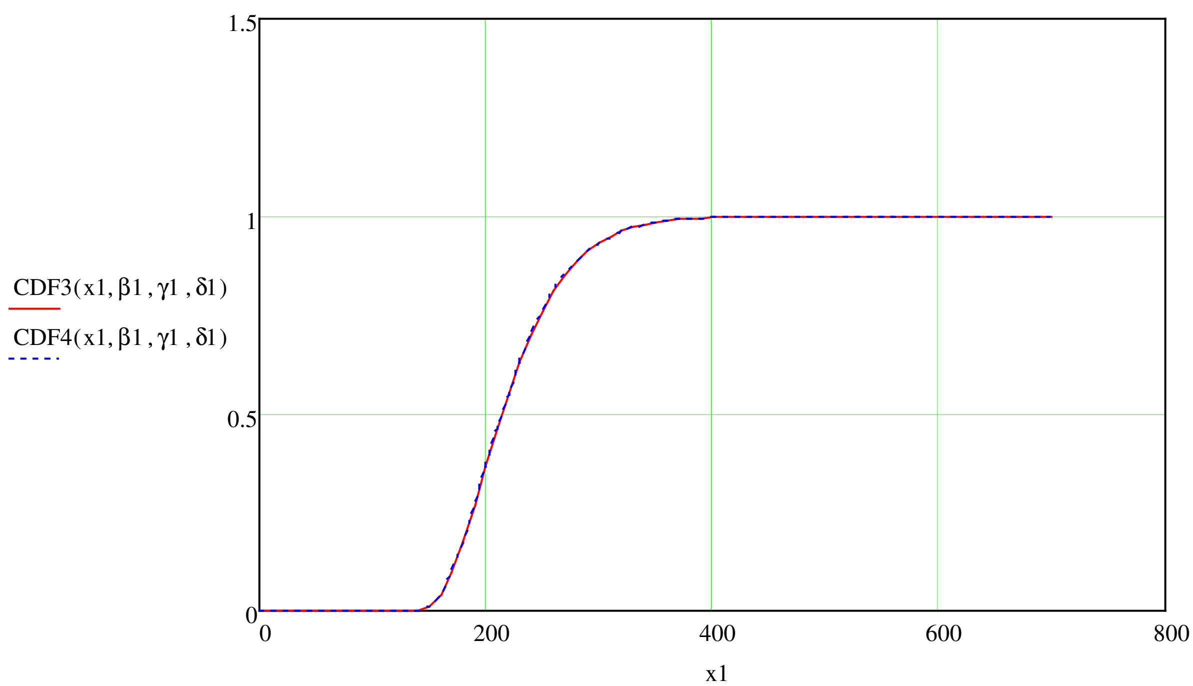
Now we set up for estimation of the three parameters beta, gamma, and delta

Here, from above, is equation that we need to fit the data to:

$$CDF3(x1, \beta 1, \gamma 1, \delta 1) := \int_0^{x1} PDF3(z, \beta 1, \gamma 1, \delta 1)\, dz$$

As it happens, the data is not a continuous function of x ... therefore we use the discrete version

$$CDF4(x1, \beta 1, \gamma 1, \delta 1) := \sum_{z=0}^{x1} PDF3(z, \beta 1, \gamma 1, \delta 1)$$



It seems that, when all is said and done, the two formulae are about the same.

In any case, we now move on to the estimation task.

Here we start a "solve block" to estimate and δ1 $k := 0, 1 .. n1$

$$SSE1(\beta2, \gamma2, \delta2) := \sum_k \left(y_k - CDF4(x_k, \beta2, \gamma2, \delta2)\right)^2$$

the sum of the squared errors is to be minimized

$$SSE5(\beta2, \gamma2, \delta2) := \sum_k \left|y_k - CDF4(x_k, \beta2, \gamma2, \delta2)\right|$$

 = scale = shape δ = offset

Given

$$SSE5(\beta2, \gamma2, \delta2) = 0$$

$\beta2 := 19.8$ $\gamma2 := 8$ $\delta2 := 267$ initial guess for parameter values

$10 < \beta2 < 30$ $3 < \gamma2 < 10$ $200 < \delta2 < 320$

$$\begin{pmatrix} \beta1est \\ \gamma1est \\ \delta1est \end{pmatrix} := Minerr(\beta2, \gamma2, \delta2)$$

$\beta1est = 19.882$ $\gamma1est = 8.027$ $\delta1est = 267.491$ Resultant estimated parameter values

$$PDF3(x, \beta1est, \gamma1est, \delta1est) := PDF1(x, \beta1est, \gamma1est, \delta1est) \cdot \Phi(x - \delta1est)$$

$\beta1est = 19.882$

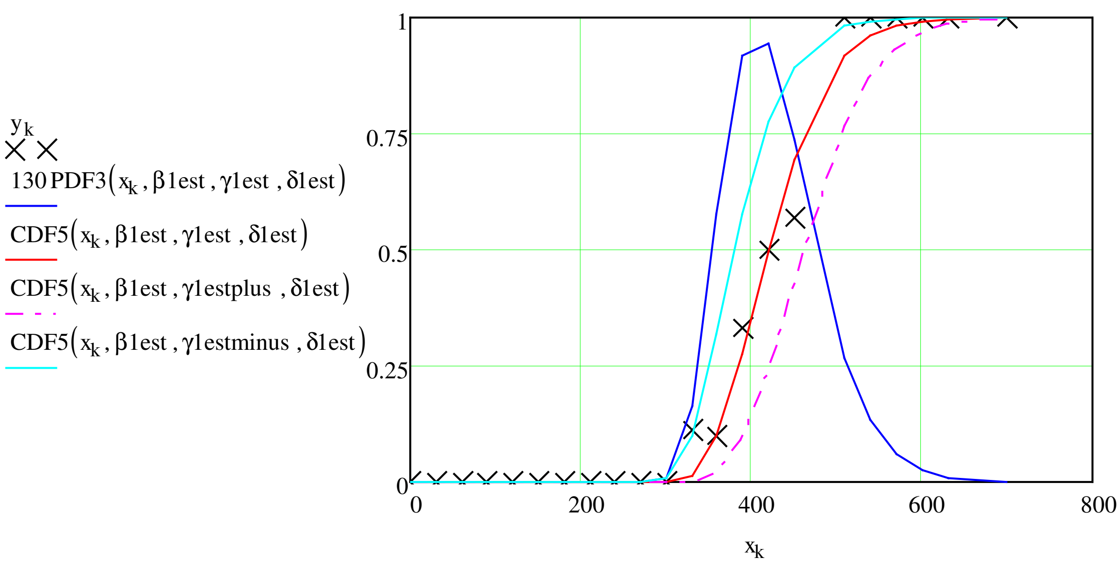
$$CDF5(x1, \beta1est, \gamma1est, \delta1est) := \int_0^{x1} PDF3(z, \beta1est, \gamma1est, \delta1est)\, dz$$

$\gamma1est = 8.027$

$$CDF4(x1, \beta1est, \gamma1est, \delta1est) := \sum_{z=0}^{x1} PDF2(z, \beta1est, \gamma1est, \delta1est)$$

$\delta1est = 267.491$

$\gamma1estplus := 1.25 \cdot \gamma1est$ $\gamma1estminus := .75 \cdot \gamma1est$ $\gamma1estplus = 10.033$ $\gamma1estminus = 6.02$



Now that we have basic curve fitting going we can calculate various parameters of the fitted curves.

Here are the first four moments of the fitted PDF3

From CRC Statistics Handbook p 77and 96

First moment - Mean

$$M1 := \int_{\delta 1est}^{\infty} x \cdot PDF3(x, \beta 1est, \gamma 1est, \delta 1est)\, dx$$

$$M1 = 427.076$$

$$m1 := \gamma 1est \cdot \beta 1est + \delta 1est$$

$$m1 = 427.076$$

Second moment - Variance

$$M2 := \int_{\delta 1est}^{\infty} (x - M1)^2 \cdot PDF3(x, \beta 1est, \gamma 1est, \delta 1est)\, dx$$

$$M2 = 3.173 \times 10^3$$

$$m2 := \gamma 1est \cdot \beta 1est^2$$

$$m2 = 3.173 \times 10^3$$

Standard Deviation

$$StdDev := \sqrt{M2} \qquad StdDev = 56.328$$

$$Mode := \beta 1est \cdot (\gamma 1est - 1) + \delta 1est$$

$$Mode = 407.194$$

Third moment - Skew

$$CoefVariation := \frac{1}{\sqrt{\gamma 1est}} \qquad CoefVariation = 0.353$$

$$M3 := \int_{\delta 1est}^{\infty} (x - M1)^3 \cdot PDF3(x, \beta 1est, \gamma 1est, \delta 1est) \cdot \left(\frac{1}{StdDev}\right)^3 dx$$

$$M3 = 0.706$$

$$m3 := \frac{2}{\sqrt{\gamma 1est}}$$

$$m3 = 0.706$$

Fourth moment - Kurtosis

$$M4 := \int_{\delta 1est}^{\infty} (x - M1)^4 \cdot PDF3(x, \beta 1est, \gamma 1est, \delta 1est) \cdot \left(\frac{1}{StdDev}\right)^4 dx$$

$$M4 = 3.748$$

$$m4 := 3\left(1 + \frac{2}{\gamma 1est}\right)$$

$$m4 = 3.748$$

Some Quartiles:

$$\int_0^{177} PDF3(z, \beta 1est, \gamma 1est, \delta 1est)\, dz = 0 \qquad \int_0^{188} PDF3(z, \beta 1est, \gamma 1est, \delta 1est)\, dz = 0$$

$$\int_0^{211} PDF3(z, \beta 1est, \gamma 1est, \delta 1est)\, dz = 0 \qquad \int_0^{243} PDF3(z, \beta 1est, \gamma 1est, \delta 1est)\, dz = 0$$

$$\int_0^{285} PDF3(z, \beta 1est, \gamma 1est, \delta 1est)\, dz = 3.875 \times 10^{-6} \qquad \int_0^{330} PDF3(z, \beta 1est, \gamma 1est, \delta 1est)\, dz = 0.015$$

Now we use the previously determined shape function to estimate the PDF from the given mean and StdDev

Shape factor = gamma = $\gamma 1est$ $\beta 1est = 19.882$ $\gamma 1est = 8.027$ $\delta 1est = 267.491$

For the gamma function we have the these relationships

$Mean\_est := \gamma 1est \cdot \beta 1est + \delta 1est$ $Variance\_est := \gamma 1est \cdot \beta 1est^2$ $StdDev\_est := \beta 1est \cdot \sqrt{\gamma 1est}$

These equations can be solved for $\beta 2$ and $\delta 2$ when the Mean, StdDev2 and shape factor $\gamma 2$ are known

$Mean\_est = 427.076$ $StdDev\_est = 56.328$

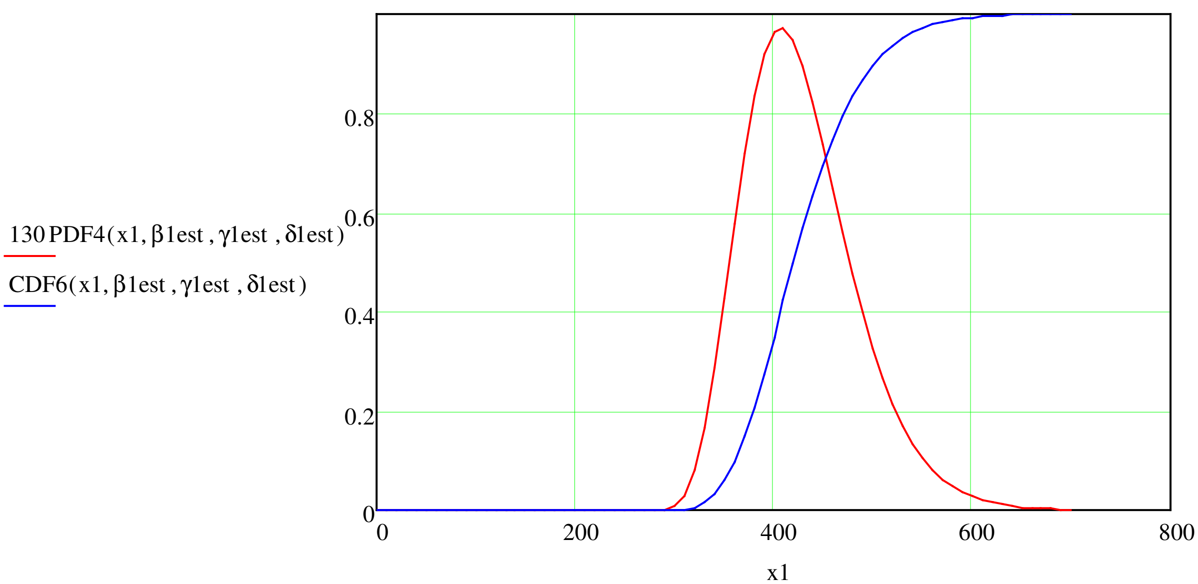
$$\beta 4 := \frac{StdDev\_est}{\sqrt{\gamma 1est}} \qquad \delta 4 := Mean\_est - StdDev\_est \cdot \left(\frac{\gamma 1est}{\sqrt{\gamma 1est}}\right)$$

$\beta 4 = 19.882$ $\delta 4 = 267.491$ $\gamma 1est = 8.027$

The resultant PDF is:

$$PDF4(x1, \beta 1est, \gamma 1est, \delta 1est) := \left[\frac{\left(\frac{x1 - \delta 1est}{\beta 1est}\right)^{\gamma 1est - 1} \cdot e^{-\frac{x1 - \delta 1est}{\beta 1est}}}{\beta 1est \cdot \Gamma(\gamma 1est)}\right] \cdot (\Phi(x1 - \delta 1est))$$

$$CDF6(x1, \beta 1est, \gamma 1est, \delta 1est) := \int_0^{x1} PDF4(z, \beta 1est, \gamma 1est, \delta 1est)\, dz$$



Guess value:

Q001 := 30(Q10 := 35(Q25 := 35(Q50 := 35(Q75 := 35(Q90 := 35(

Now we use a "Solve Block" to find all the Quantiles of interest

Quantile .001: Guess value: 200 $Q001 = 300$

Given

$$CDF6(x1, \beta 1est, \gamma 1est, \delta 1est) := \int_0^{x1} PDF4(z, \beta 1est, \gamma 1est, \delta 1est)\, dz$$

$CDF6(Q001, \beta 1est, \gamma 1est, \delta 1est) = .001$ $T001 := Find(Q001)$ $T001 = 306.922$

Quantile .1: Guess value: 250 $Q10 = 350$

Given

$$CDF6(x1, \beta 1est, \gamma 1est, \delta 1est) := \int_0^{x1} PDF4(z, \beta 1est, \gamma 1est, \delta 1est)\, dz$$

$CDF6(Q10, \beta 1est, \gamma 1est, \delta 1est) = .1$ $T10 := Find(Q10)$ $T10 = 360.471$

Quantile .25: Guess value: 250 $Q25 = 350$

Given

$$CDF6(x1, \beta 1est, \gamma 1est, \delta 1est) := \int_0^{x1} PDF4(z, \beta 1est, \gamma 1est, \delta 1est)\, dz$$

$CDF6(Q25, \beta 1est, \gamma 1est, \delta 1est) = .25$ $T25 := Find(Q25)$ $T25 = 386.375$

Quantile .5: Guess value: 275 $Q50 = 350$

Given

$$CDF6(x1, \beta 1est, \gamma 1est, \delta 1est) := \int_0^{x1} PDF4(z, \beta 1est, \gamma 1est, \delta 1est)\, dz$$

$CDF6(Q50, \beta 1est, \gamma 1est, \delta 1est) = .5$ $T50 := Find(Q50)$ $T50 = 420.5$

Quantile .75: Guess value: 300 $Q75 = 350$

Given

$$CDF6(x1, \beta 1est, \gamma 1est, \delta 1est) := \int_0^{x1} PDF4(z, \beta 1est, \gamma 1est, \delta 1est)\, dz$$

$CDF6(Q75, \beta 1est, \gamma 1est, \delta 1est) = .75$ $T75 := Find(Q75)$ $T75 = 460.631$

Quantile .9: Guess value: 400 $Q90 = 350$

Given

$$CDF6(x1, \beta 1est, \gamma 1est, \delta 1est) := \int_0^{x1} PDF4(z, \beta 1est, \gamma 1est, \delta 1est)\, dz$$

$CDF6(Q90, \beta 1est, \gamma 1est, \delta 1est) = .9$ $T90 := Find(Q90)$ $T90 = 502.172$