Mean, Variance, and Simulation of the Truncated Normal Distribution

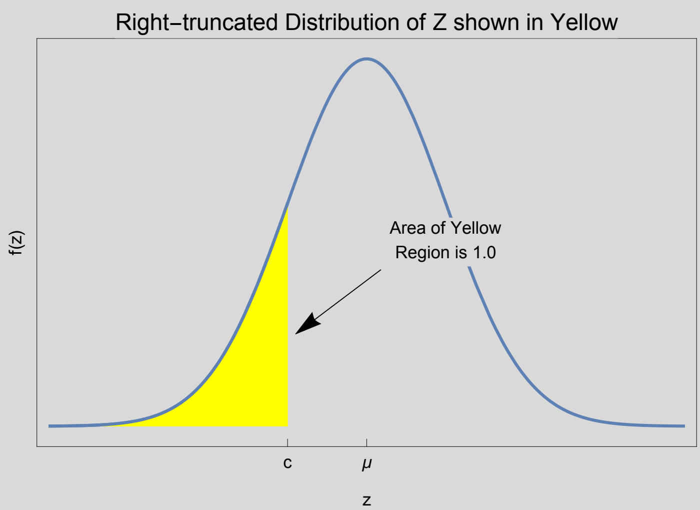
Nagham Mohammad, Matthew McLeod, and Ian McLeod

Department of Statistical and Actuarial Sciences
Western University
London, Ontario, Canada
August 17, 2014

Truncated and censored normal distribution

Let X be normally distributed with mean μ and variance σ^2 and let c be a fixed constant. Let Z be defined as the condtional distribution of X given $X < c$. Then Z has a right-truncated normal distribution with parameters μ, σ^2, and c. The probability density function for Z is $\phi(x)/\Phi(c)$, where $\phi(x)$ and $\Phi(c)$ denote the probability density and cumulative distribution function for X.

```
SetDirectory[NotebookDirectory[]];
g1 = Plot[PDF[NormalDistribution[0, 1], z], {z, -4, 4}, Axes → False,
    Frame → True, FrameTicks → {{None, None}, {{{0, "μ"}, {-1, "c"}}, None}},
    FrameLabel → {"z", "f(z)"}, PlotLabel →
     "Right-truncated Distribution of Z shown in Yellow",
    Epilog → {Text["Area of Yellow\nRegion is 1.0", {1, 0.2}],
      Arrow[{{0.18, 0.17}, {-0.9, 0.1}}]}];
g2 = Plot[PDF[NormalDistribution[0, 1], z], {z, -4, -1}, Axes → False,
    Frame → True, FrameTicks → {{None, None}, {{{0, "μ"}, {-1, "c"}}, None}},
    Filling → Axis, FillingStyle → Yellow];
Show[
 {g1,
  g2}]
```



If $Y = \max(X, c)$, we say Y is left-censored at c. Right-truncated distributions arise in left-censoring, when censoring occurs, the unobserved latent variable Z has a right-truncated distribution.

Note that the distribution of Y is not left-truncated since it is a mixed distribution with $\Pr\{Y = c\} > 0$. However the conditional distribution of Y greater than c is left-truncated.

Similarly the conditional distribution of $X > c$ is said to the left-truncated and it corresponds to the distribution of the latent variable in the case of right-censoring.

The mean and variance of truncated normal distributions were discussed by Barr and Sherrill (1999) but with *Mathematica* Version 10 it is easy to compute symbolic formula for these quantities. The mvstn package was created by using *Mathematica* to compute the mean and variance symbolically and then using the *Mathematica* function CForm[] to convert to an expression in C. This C code was then used to create C functions which were interfaced to R in the package mvstn.

Reference: Donald R. Barr and E. Todd Sherrill (1999). Mean and Variance of Truncated Normal Distributions. The American Statistician, Vol. 53, No. 4 (Nov., 1999), pp. 357-361.

Mean and variance in right truncated normal

C code generation

We compute the mean symbolically, show the C code, timings, and then export it to a file.

Mean

```
Timing[MeanZR = N[Simplify[
    Mean[TruncatedDistribution[{-∞, c}, NormalDistribution[zmu, zsig]]],
    Assumptions → zmu ∈ Reals && zsig > 0.0]]]
```

$$\left\{5.703125, \frac{\text{zmu} - 0.797885 \times 2.71828^{-\frac{0.5\,(c-1.\,\text{zmu})^2}{\text{zsig}^2}}\,\text{zsig} + \text{zmu}\,\text{Erf}\left[\frac{0.707107\,(c-1.\,\text{zmu})}{\text{zsig}}\right]}{\text{Erfc}\left[\frac{0.707107\,(-1.\,c+\text{zmu})}{\text{zsig}}\right]}\right\}$$

```
CForm[MeanZR]
```

```
(zmu - (0.7978845608028654*zsig)/
      Power(2.718281828459045,(0.5*Power(c - 1.*zmu,2))/Power(zsig,2)) +
     zmu*Erf((0.7071067811865475*(c - 1.*zmu))/zsig))/
   Erfc((0.7071067811865475*(-1.*c + zmu))/zsig)
```

```
SetDirectory[NotebookDirectory[]];
WriteString["MeanZR.c", Evaluate[CForm[MeanZR]]]
```

Variance

```
Timing[varZR = N[FullSimplify[
    Variance[TruncatedDistribution[{-∞, c}, NormalDistribution[zmu, zsig]]],
    Assumptions → zmu ∈ Reals && zsig > 0.0]]]
```

$$\Big\{57.984375,\ \frac{1}{\text{Erfc}\left[\frac{0.707107\,(-1.\,c+\text{zmu})}{\text{zsig}}\right]^{3}}\,0.225079\times 2.71828^{-\frac{1.\,(c^{2}+\text{zmu}^{2})}{\text{zsig}^{2}}}\ \text{zsig}$$
$$\left(1.+\text{Erf}\left[\frac{0.707107\,(c-1.\,\text{zmu})}{\text{zsig}}\right]\right)\left(1.41421\ \text{zsig}\left(-2.\ 2.71828^{\frac{2.\,c\,\text{zmu}}{\text{zsig}^{2}}}+\right.\right.$$
$$3.14159\times 2.71828^{\frac{c^{2}+\text{zmu}^{2}}{\text{zsig}^{2}}}\left(1.+\text{Erf}\left[\frac{0.707107\,(c-1.\,\text{zmu})}{\text{zsig}}\right]\right)^{2}+$$
$$\left.\left.3.54491\times 2.71828^{\frac{0.5\,(c+\text{zmu})^{2}}{\text{zsig}^{2}}}\,(c-1.\,\text{zmu})\left(-2.+\text{Erfc}\left[\frac{0.707107\,(c-1.\,\text{zmu})}{\text{zsig}}\right]\right)\right)\right)\Big\}$$

```
CForm[varZR]
```

```
(0.22507907903927651*zsig*(1. + Erf((0.7071067811865475*(c - 1.*zmu))/zsig))*
     (1.4142135623730951*zsig*(-2.*Power(2.718281828459045,(2.*c*zmu)/Power(zsi
          3.141592653589793*Power(2.718281828459045,
            (Power(c,2) + Power(zmu,2))/Power(zsig,2))*
           Power(1. + Erf((0.7071067811865475*(c - 1.*zmu))/zsig),2)) +
        3.5449077018110318*Power(2.718281828459045,(0.5*Power(c + zmu,2))/Power(
         (c - 1.*zmu)*(-2. + Erfc((0.7071067811865475*(c - 1.*zmu))/zsig)))))/
   (Power(2.718281828459045,(1.*(Power(c,2) + Power(zmu,2)))/Power(zsig,2))*
     Power(Erfc((0.7071067811865475*(-1.*c + zmu))/zsig),3))
```

```
WriteString["VarZR.c", Evaluate[CForm[varZR]]]
```

Mathematica functions and testing

```
meanzr[zmu_, zsig_, c_] := Evaluate[MeanZR]
```

```
varzr[zmu_, zsig_, c_] := Evaluate[varZR]
```

```
meanzr[100.0, 15.0, 80.0]
```

```
73.028
```

```
varzr[100.0, 15.0, 80.0]
```

```
36.9504
```

```
Sqrt[%]
```

```
6.07869
```

Mean and variance in left truncated normal

C code generation

We compute the mean symbolically, show the C code, and then export it to a file.

Mean

```
Timing[MeanZL = N[FullSimplify[
    Mean[TruncatedDistribution[{c, ∞}, NormalDistribution[zmu, zsig]]]]]]
```

$$\left\{5.703125, \frac{2.71828^{-\frac{0.5\,(c-1.\,zmu)^2}{zsig^2}}\left(0.797885\,zsig + 2.71828^{\frac{0.5\,(c-1.\,zmu)^2}{zsig^2}}\,zmu\,\mathrm{Erfc}\left[\frac{0.707107\,(c-1.\,zmu)}{zsig}\right]\right)}{1. + \mathrm{Erf}\left[\frac{0.707107\,(-1.\,c+zmu)}{zsig}\right]}\right\}$$

```
CForm[MeanZL]
```

```
(0.7978845608028654*zsig + Power(2.718281828459045,
      (0.5*Power(c - 1.*zmu,2))/Power(zsig,2))*zmu*
     Erfc((0.7071067811865475*(c - 1.*zmu))/zsig))/
  (Power(2.718281828459045,(0.5*Power(c - 1.*zmu,2))/Power(zsig,2))*
    (1. + Erf((0.7071067811865475*(-1.*c + zmu))/zsig)))
```

```
SetDirectory[NotebookDirectory[]];
WriteString["MeanZL.c", Evaluate[CForm[MeanZL]]]
```

Variance

```
Timing[varZL = N[FullSimplify[
    Variance[TruncatedDistribution[{c, ∞}, NormalDistribution[zmu, zsig]]],
    Assumptions → zmu ∈ Reals && zsig > 0]]]
```

$$\left\{40.875000, \left(0.225079 \times 2.71828^{-\frac{1.\,(c^2+zmu^2)}{zsig^2}}\,zsig\,\mathrm{Erfc}\left[\frac{0.707107\,(c-1.\,zmu)}{zsig}\right]\right.\right.$$
$$\left(3.54491 \times 2.71828^{\frac{0.5\,(c+zmu)^2}{zsig^2}}\,(c-1.\,zmu)\,\mathrm{Erfc}\left[\frac{0.707107\,(c-1.\,zmu)}{zsig}\right] - \right.$$
$$1.41421\,zsig\left(2. \times 2.71828^{\frac{2.\,c\,zmu}{zsig^2}} - 3.14159 \times 2.71828^{\frac{c^2+zmu^2}{zsig^2}}\right.$$
$$\left.\left.\left.\left.\mathrm{Erfc}\left[\frac{0.707107\,(c-1.\,zmu)}{zsig}\right]^2\right)\right)\right) \Big/ \left(1. + \mathrm{Erf}\left[\frac{0.707107\,(-1.\,c+zmu)}{zsig}\right]\right)^3\right\}$$

```
CForm[varZL]
```

```
(0.22507907903927651*zsig*Erfc((0.7071067811865475*(c - 1.*zmu))/zsig)*
     (3.5449077018110318*Power(2.718281828459045,(0.5*Power(c + zmu,2))/Power(z
        (c - 1.*zmu)*Erfc((0.7071067811865475*(c - 1.*zmu))/zsig) -
       1.4142135623730951*zsig*(2.*Power(2.718281828459045,(2.*c*zmu)/Power(zsi
          3.141592653589793*Power(2.718281828459045,
            (Power(c,2) + Power(zmu,2))/Power(zsig,2))*
           Power(Erfc((0.7071067811865475*(c - 1.*zmu))/zsig),2))))/
    (Power(2.718281828459045,(1.*(Power(c,2) + Power(zmu,2)))/Power(zsig,2))*
      Power(1. + Erf((0.7071067811865475*(-1.*c + zmu))/zsig),3))
```

```
WriteString["VarZL.c", Evaluate[CForm[varZL]]]
```

Mathematica functions and testing

```
meanzl[zmu_, zsig_, c_] := Evaluate[MeanZL]
```

```
varzl[zmu_, zsig_, c_] := Evaluate[varZL]
```

```
meanzl[100.0, 15.0, 80.0]
```

```
102.707
```

```
varzl[100.0, 15.0, 80.0]
```

```
163.53
```

```
Sqrt[%]
```

```
12.7879
```

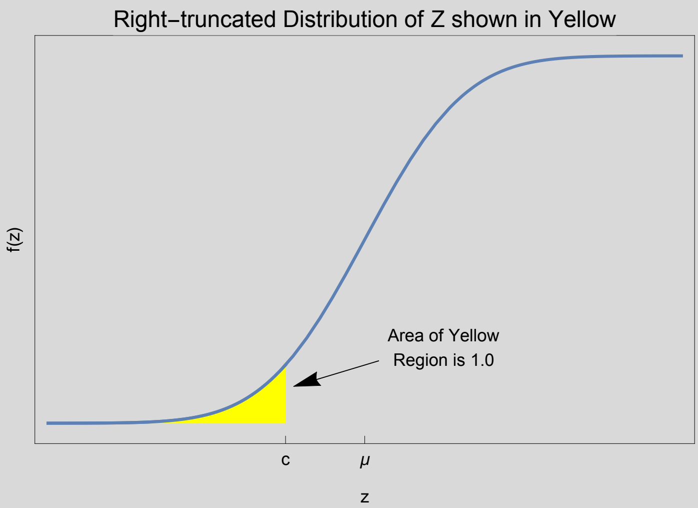
Simulate Truncated Normal Distribution

Right truncated case

In[57]:=

```
SetDirectory[NotebookDirectory[]];
g1 = Plot[CDF[NormalDistribution[0, 1], z], {z, -4, 4}, Axes → False,
   Frame → True, FrameTicks → {{None, None}, {{{0, "μ"}, {-1, "c"}}, None}},
   FrameLabel → {"z", "f(z)"}, PlotLabel →
    "Right-truncated Distribution of Z shown in Yellow",
   Epilog → {Text["Area of Yellow\nRegion is 1.0", {1, 0.2}],
     Arrow[{{0.18, 0.17}, {-0.9, 0.1}}]}];
g2 = Plot[CDF[NormalDistribution[0, 1], z], {z, -4, -1}, Axes → False,
   Frame → True, FrameTicks → {{None, None}, {{{0, "μ"}, {-1, "c"}}, None}},
   Filling → Axis, FillingStyle → Yellow];
Show[
 {g1,
  g2}]
```

Out[60]=




```
SimulateRTN[n_, μ_, σ_, c_] := Module[{},
  U = RandomVariate[UniformDistribution[{0, 1}], n];
  Quantile[NormalDistribution[μ, σ], U * CDF[NormalDistribution[μ, σ], c]]
 ]
```

```
z = SimulateRTN[10^6, 100, 15, 80];
{Mean[#], Variance[#]} &[z]
```

```
{73.0289, 36.8494}
```

```
z = SimulateRTN[10^6, 100, 15, 100];
{Mean[#], Variance[#]} &[z]
```

```
{88.0408, 81.6746}
```

```
z = SimulateRTN[10^6, 100, 15, 110];
{Mean[#], Variance[#]} &[z]
```

```
{93.5819, 120.06}
```

Validation

```
> mvtn(100,15,80,"right")
[1] 73.02798 36.95043
> mvtn(100,15,100,"right")
[1] 88.03173 81.76055
> mvtn(100,15,110,"right")
[1]  93.58974 119.80588
```

Left truncated case

Due to symmetry, if Z has a right-truncated distribution with truncation point c, and parameters (μ, σ) then $-Z$ has a left-truncated distribution with parameters $(-\mu, \sigma)$ and truncation point $-c$. So to simulate from a left truncated distribution with parameters (μ, σ, c) we can simulate from a right truncated distribution with parameters $(-\mu, \sigma, -c)$ and negate the result.

Simple method

```
SimulateLTN[n_, μ_, σ_, c_] := -SimulateRTN[n, -μ, σ, -c];
```

```
z = SimulateLTN[10^6, 100, 15, 80];
{Mean[#], Variance[#]} &[z]
```

```
{102.705, 163.687}
```

```
z = SimulateLTN[10^6, 100, 15, 100];
{Mean[#], Variance[#]} &[z]
```

```
{111.969, 81.9788}
```

```
z = SimulateLTN[10^6, 100, 15, 110];
{Mean[#], Variance[#]} &[z]
```

```
{118.964, 54.5512}
```

```
> mvtn(100,15,80,"left")
[1] 102.7071 163.5305
> mvtn(100,15,100,"left")
[1] 111.96827  81.76055
> mvtn(100,15,110,"left")
[1] 118.97767  54.62474
```