Random Variables as a Function of $Uniform(a,b)$

Let $X = 1/V$, V a $Uniform(a,b)$ $0 < a < b$ variable. Find the supporting set $\mathcal{X}$, cdf and idf ($F(x)$, $F^{-1}(u)$).

X is a continuous **decreasing** function of V, given $0 < a < b$ the supporting set is therefore $\mathcal{X} = (\frac{1}{b}, \frac{1}{a})$.

From the definition of cdf:

$$F(x) = Pr(X \leq x) = Pr\left(\frac{1}{V} \leq x\right) = Pr\left(V \geq \frac{1}{x}\right)$$

$$F(x) = 1 - \left(\frac{\frac{1}{x} - a}{b-a}\right) \quad \text{by defns of probability and } V \text{ a } Uniform$$

$$F(x) = \frac{b - a - \frac{1}{x} + a}{b-a} = \frac{b - \frac{1}{x}}{b-a} \quad \text{check the endpoints of } \mathcal{X} \in \left(\frac{1}{b}, \frac{1}{a}\right) !$$

Random Variables as a Function of $Uniform(a,b)$

Let $X = 1/V$, V a $Uniform(a,b)$ $0 < a < b$ variable. Find the supporting set $\mathcal{X}$, cdf and idf ($F(x)$, $F^{-1}(u)$).

To find the idf, $F^{-1}(u)$ (an algebraic variate) set $F(x) = u$ and solve for x:

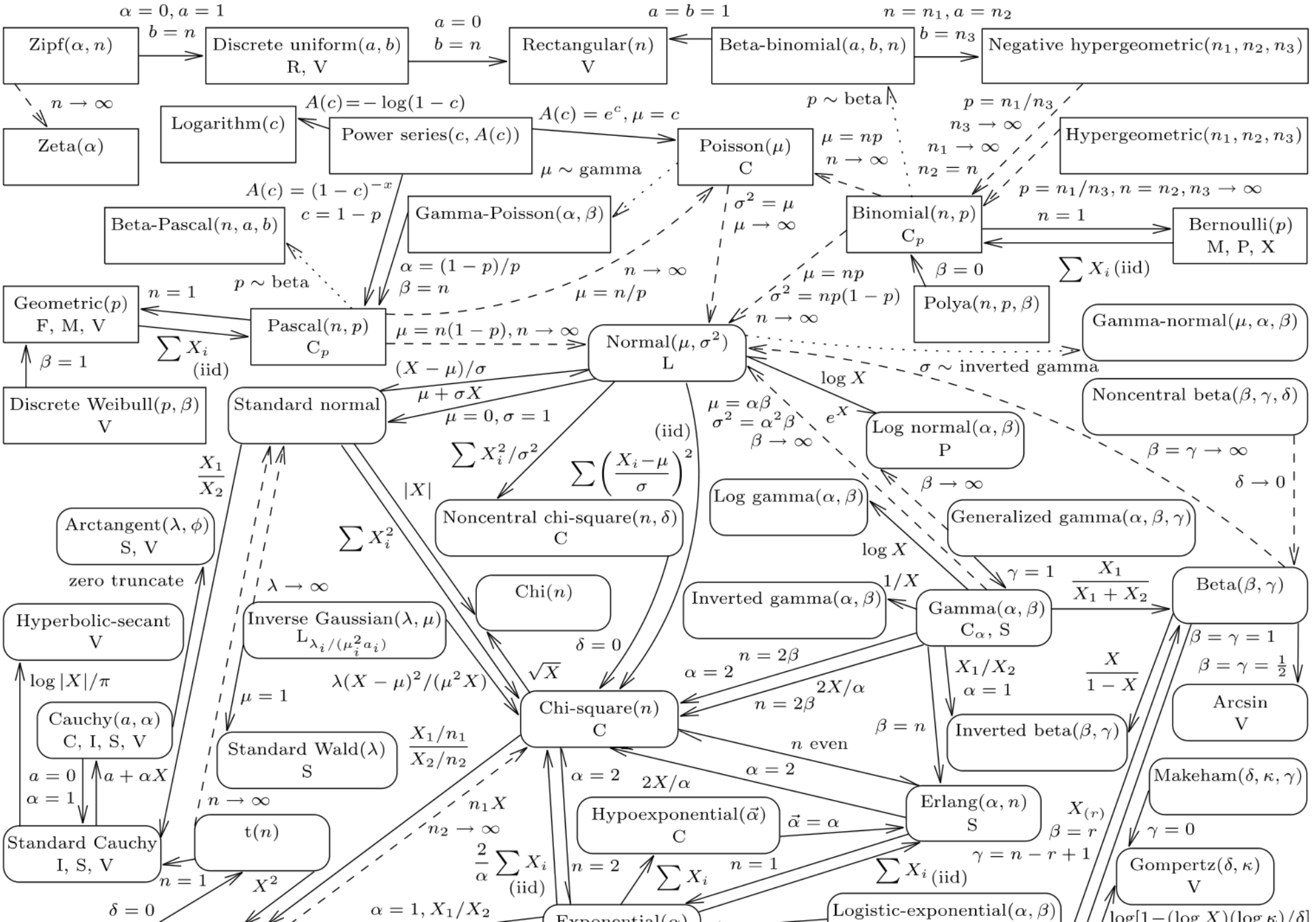
$$\begin{aligned} u &= \frac{b - \frac{1}{x}}{b-a} \\ x &= \frac{1}{b - u(b-a)} \end{aligned}$$

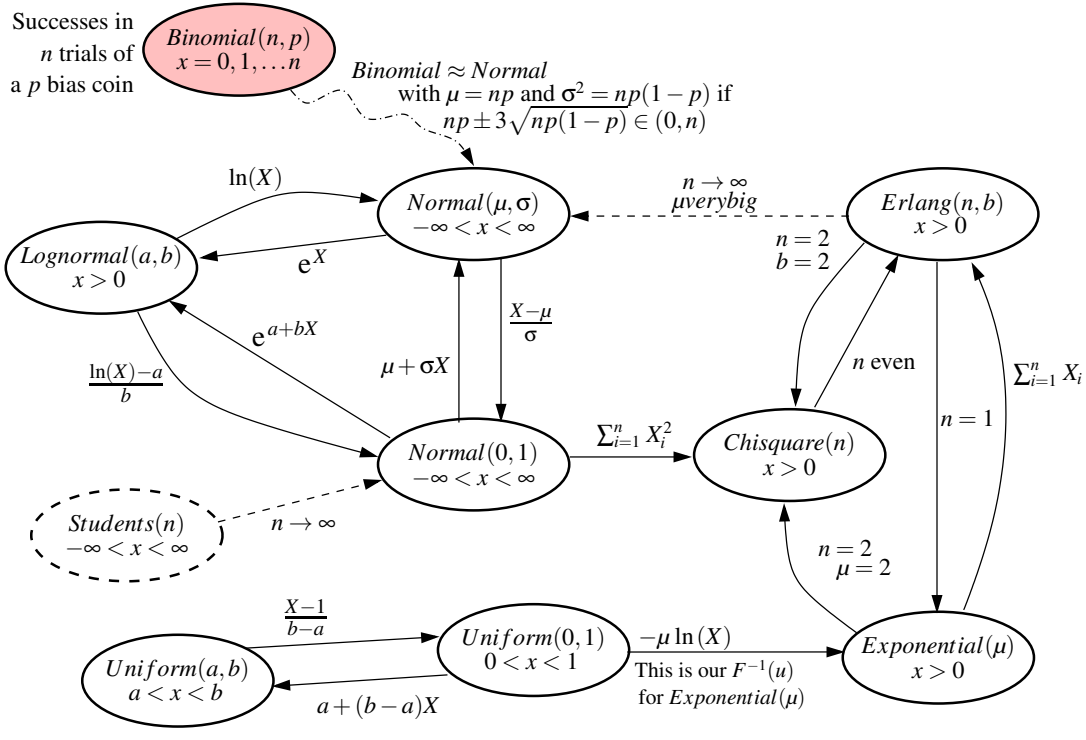
check the endpoints generated by $u \in (0,1)$!

Continuous Random Variate Generation

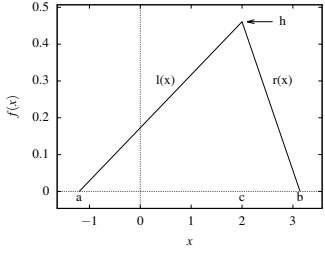
i. Inversion (solve $x = F^{-1}(u)$ algebraically).
ii. Approximate Inversion (solve $x = H(u)$, where $H(u) \approx F^{-1}(u)$)
iii. Numerical Inversion (solve $x = F^{-1}(u)$ numerically).
iv. Construction (as in DRVs)
v. Acceptance-Rejection with a *majoring* $g(\cdot)$ and known $G^{-1}(u)$

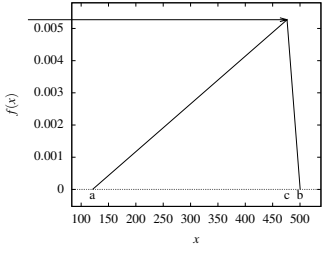
We still use: portability, exactness, robustness, efficiency, clarity, syncronization, and monotonicity as metrics to evaluate random variates of CRVs.

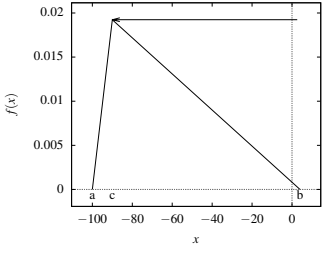

Zipf(α, n)
Discrete uniform(a, b)
R, V
Rectangular(n)
V
Beta-binomial(a, b, n)
Negative hypergeometric(n1, n2, n3)
Zeta(α)
Logarithm(c)
Power series(c, A(c))
Poisson(μ)
C
Hypergeometric(n1, n2, n3)
Beta-Pascal(n, a, b)
Gamma-Poisson(α, β)
Binomial(n, p)
Cp
Bernoulli(p)
M, P, X
Geometric(p)
F, M, V
Pascal(n, p)
Cp
Polya(n, p, β)
Gamma-normal(μ, α, β)
Normal(μ, σ2)
L
Discrete Weibull(p, β)
V
Standard normal
Noncentral beta(β, γ, δ)
Log normal(α, β)
P
Log gamma(α, β)
Generalized gamma(α, β, γ)
Arctangent(λ, φ)
S, V
Noncentral chi-square(n, δ)
C
Chi(n)
Inverted gamma(α, β)
Gamma(α, β)
Cα, S
Beta(β, γ)
Hyperbolic-secant
V
Inverse Gaussian(λ, μ)
Chi-square(n)
C
Arcsin
V
Cauchy(a, α)
C, I, S, V
Standard Wald(λ)
S
Inverted beta(β, γ)
Makeham(δ, κ, γ)
Erlang(α, n)
S
Standard Cauchy
I, S, V
t(n)
Hypoexponential(α⃗)
C
Gompertz(δ, κ)
V
Logistic-exponential(α, β)
Exponential(α)
zero truncate
α = 0, a = 1
b = n
a = 0
b = n
a = b = 1
n = n1, a = n2
b = n3
n → ∞
A(c) = − log(1 − c)
A(c) = e^c, μ = c
p ∼ beta
p = n1/n3
n3 → ∞
n1 → ∞
n2 = n
μ = np
n → ∞
μ ∼ gamma
A(c) = (1 − c)^−x
c = 1 − p
σ2 = μ
μ → ∞
p = n1/n3, n = n2, n3 → ∞
n = 1
∑ Xi (iid)
α = (1 − p)/p
β = n
μ = n/p
β = 0
μ = np
σ2 = np(1 − p)
β = 1
μ = n(1 − p), n → ∞
(X − μ)/σ
μ + σX
μ = 0, σ = 1
σ ∼ inverted gamma
log X
μ = αβ
σ2 = α2β
β → ∞
e^X
β = γ → ∞
δ → 0
X1/X2
(iid)
∑ Xi2/σ2
∑ ((Xi − μ)/σ)^2
|X|
∑ Xi2
λ → ∞
1/X
γ = 1
X1/(X1 + X2)
μ = 1
λ(X − μ)2/(μ2X)
√X
δ = 0
n = 2β
α = 2
2X/α
X1/X2
α = 1
X/(1 − X)
β = γ = 1
β = γ = 1/2
β = n
n even
a = 0
α = 1
a + αX
X1/n1
X2/n2
n → ∞
n1X
n2 → ∞
2/α ∑ Xi (iid)
X(r)
β = r
γ = n − r + 1
γ = 0
δ = 0
X2
α = 1, X1/X2
log |X|/π


Successes in n trials of a p bias coin
$Binomial(n,p)$ $x = 0,1,\ldots n$
$Binomial \approx Normal$ with $\mu = np$ and $\sigma^2 = np(1-p)$ if $np \pm 3\sqrt{np(1-p)} \in (0,n)$
$\ln(X)$
$Normal(\mu,\sigma)$ $-\infty < x < \infty$
$n \to \infty$ $\mu verybig$
$Erlang(n,b)$ $x > 0$
$Lognormal(a,b)$ $x > 0$
e^X
$n = 2$ $b = 2$
e^{a+bX}
$\frac{X-\mu}{\sigma}$
$\frac{\ln(X)-a}{b}$
$\mu + \sigma X$
n even
$\sum_{i=1}^{n} X_i$
$n = 1$
$\sum_{i=1}^{n} X_i^2$
$Chisquare(n)$ $x > 0$
$Normal(0,1)$ $-\infty < x < \infty$
$Students(n)$ $-\infty < x < \infty$
$n \to \infty$
$n = 2$ $\mu = 2$
$\frac{X-1}{b-a}$
$Uniform(0,1)$ $0 < x < 1$
$-\mu \ln(X)$
This is our $F^{-1}(u)$ for $Exponential(\mu)$
$Exponential(\mu)$ $x > 0$
$Uniform(a,b)$ $a < x < b$
$a + (b-a)X$

The Triangular Distribution

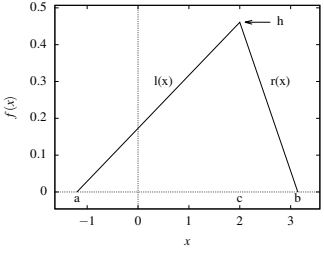






1. Useful in the **specification model** when you know (or have a good guess for) only the **minimum**, **mode** (most likely), and **maximum** of a variable x's distribution.
2. a, c, b seem awkward, but recall that a and b are the canonical **endpoints** for finite distributions ($Uniform(a,b)$, $(Equilikely(a,b), \ldots)$.
3. Author cites scheduling simulations for shuttle missions, apparently terrestrial practice and rehearsals don't always translate to zero gravity environments.

$F^{-1}(u)$ by Inversion — Roadmap for Triangular Distribution

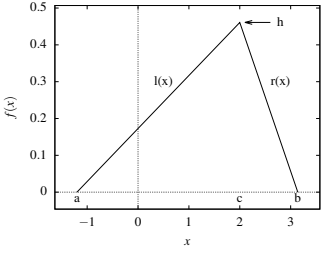


1. Given $a < c < b$ and the area under the $l(x)$ and $r(x)$ is 1, h can be calculated.
2. With h, lines $l(x)$ and $r(x)$ can be calculated and we know a (piecewise) $f(x)$

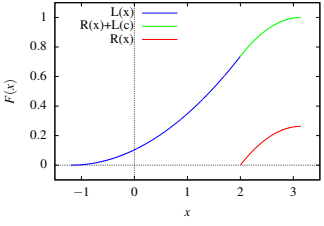
$$f(x) = \begin{cases} a < x \leq c & l(x) \\ c < x < b & r(x) \end{cases}$$

Next, we need to find the piecewise CDF $F(x)$ of $f(x)\ldots$

$F^{-1}(u)$ by Inversion — Roadmap for Triangular Distribution



Piecewise CDF Triangular Distribution

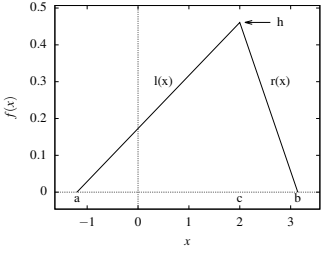


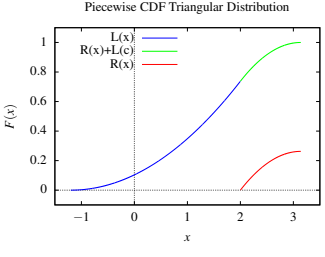
3. Find $L(x)$ and $R(x)$, together they form $F(x)\ldots$

$$L(x) = \int_a^x l(t)\,\mathrm{d}t \quad R(x) = \int_c^x r(t)\,\mathrm{d}t$$

$$F(x) = \begin{cases} a < x \le c & L(x) \\ c < x < b & L(c) + R(x) \end{cases}$$

$F^{-1}(u)$ by Inversion — Roadmap for Triangular Distribution





3. Find $L(x)$ and $R(x)$, together they form $F(x)\ldots$

$$L(x) = \int_a^x l(t)\,\mathrm{d}t \quad R(x) = \int_c^x r(t)\,\mathrm{d}t$$

$$F(x) = \begin{cases} a < x \le c & L(x) \\ c < x < b & L(c) + R(x) \end{cases}$$

4. *The quadratic equation is your friend!* Solve $u = L(x)$ and $u = L(c) + R(x)$ for x

$$x = \begin{cases} 0 < u \le L(c) & L^{-1}(u) \\ L(c) < u < 1 & R^{-1}(u - L(c)) \end{cases}$$

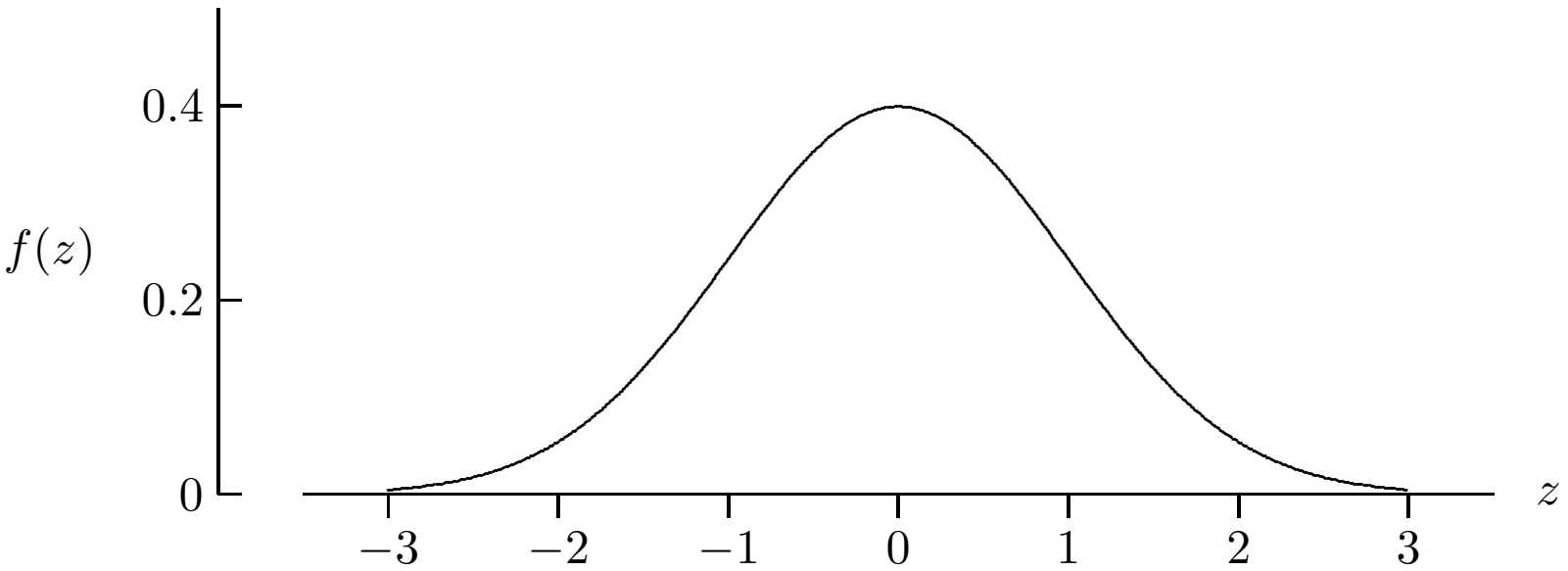
And this is our $Triangle(a,c,b)$ **analytical variate by inversion**.

Continuous Random Variable Models

- **Standard Normal Random Variable**

  Z is $Normal(0, 1)$ if and only if the set of all possible values is $\mathcal{Z} = (-\infty, \infty)$ and the pdf is

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \qquad -\infty < z < \infty$$



Standard Normal Random Variable

- If Z is $Normal(0,1)$, Z is "standardized"

  The mean is

$$\mu = \int_{-\infty}^{\infty} zf(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \exp(-z^2/2)dz = \cdots = 0$$

  The variance is

$$\sigma^2 = \int_{-\infty}^{\infty} (z-\mu)^2 f(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp(-z^2/2)dz = \cdots = 1$$

- The cdf is

$$F(z) = \int_{-\infty}^{z} f(t)dt = \Phi(z) \qquad -\infty < z < \infty$$

Standard Normal cdf

- $\Phi(\cdot)$ is defined as

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)dt \qquad -\infty < z < \infty$$

- No closed-form expression for $\Phi(z)$

$$\Phi(z) = \begin{cases} \dfrac{1 + P(1/2, z^2/2)}{2} & z \geq 0 \\ 1 - \Phi(z) & z < 0 \end{cases}$$

  $P(a, x)$ is an incomplete gamma function (see Appendix D)

- Function $\Phi(z)$ is available in `rvms` as `cdfNormal(0.0, 1.0, z)`

Approximate Inversion

- If Z is a $Normal(0, 1)$, the cdf is the special function $\Phi(\cdot)$
- The idf $\Phi^{-1}(\cdot)$ cannot be evaluated in closed form
- The idf can be *approximated* as the ratio of two fourth degree polynomials (Odeh and Evans, 1974)
- The approximation is efficient and essentially has negligible error

Numerical Inversion

- *Numerical* inversion provides another way to generate continuous random variates; that is, $u = F(x)$ can be solved for x iteratively
- *Newton's method* provides a good compromise between rate of convergence and robustness
- Given $u \in (0, 1)$, let t be close to the value of x for which $u = F(x)$
- If $F(\cdot)$ is expanded in a Taylor's series about the point t

$$F(x) = F(t) + F'(t)(x - t) + \frac{1}{2!}F''(t)(x - t)^2 + \cdots$$

- Recall $F'(t) = f(t)$
- For small $|x - t|$, ignore $(x - t)^2$ and higher order terms

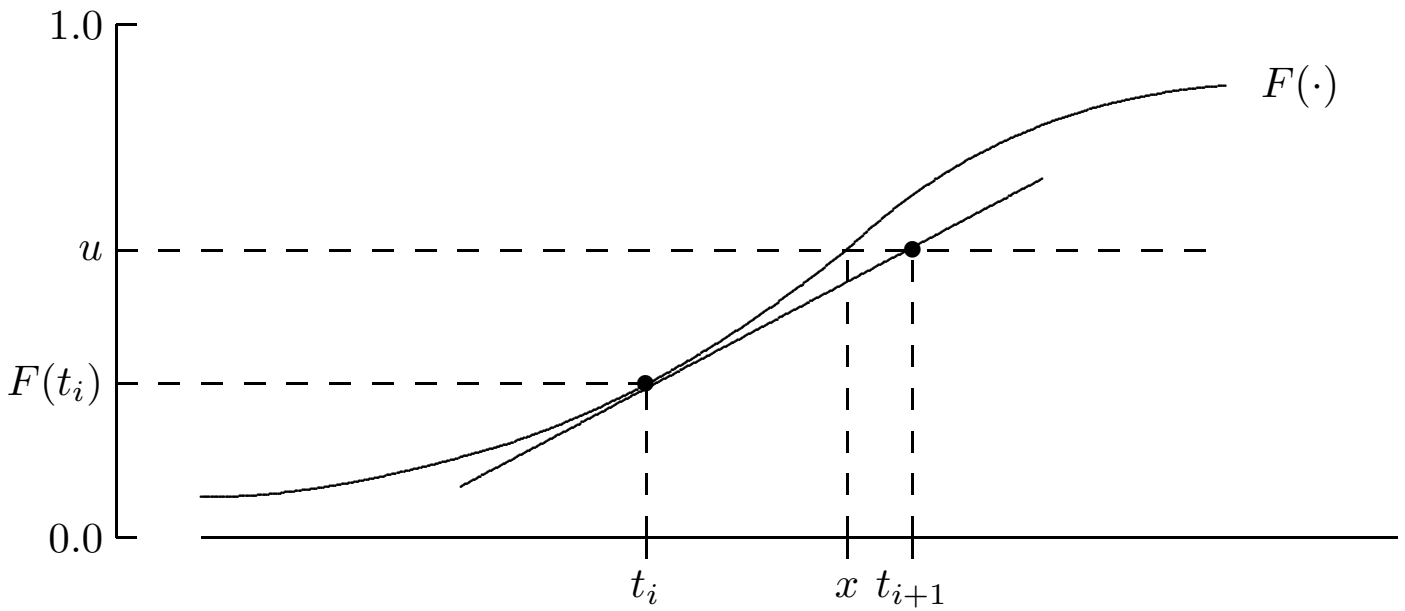
Newton's Method

- Set $u = F(x) \simeq F(t) + f(t)(x - t)$ and solve for x to obtain

$$x \simeq t + \frac{u - F(t)}{f(t)}$$

- Use initial guess t_0 and iterate to solve for x: $t_i \longrightarrow x$ as $i \longrightarrow \infty$

$$t_{i+1} = t_i + \frac{u - F(t_i)}{f(t_i)} \qquad i = 0, 1, 2, \cdots$$



Two Issues Relative to Newton's Method

- The choice of an initial value t_0
  - The best choice for the initial value is the mode
  - For most continuous RVs described in text, $t_0 = \mu$ is an essentially equivalent choice
- The test for convergence
  - Given a convergence parameter $\epsilon > 0$
  - Iterate until $|t_{i+1} - t_i| < \epsilon$

Alternative Random Variate Generation Algorithms

- **Erlang Random Variates**

  An *Erlang*(n, b) random variate can be generated by summing n *Exponential*(b) random variates

Generating an *Erlang*(n, b) Random Variate

```
x = 0.0;
for (i = 0; i < n; i++)
    x += Exponential(b);
return x;
```

- The algorithm is: portable, exact, robust, and clear
- The algorithm is **not** efficient (it is $\mathcal{O}(n)$), synchronized or monotone

Other Variates by Construction

$Chisquare(n)$ is the sum of squared $Normal(0,1)$ variables...

$x \leftarrow 0.0$
repeat n `times (`
    $X_i \leftarrow Normal(0,1)$
    $x \leftarrow x + X_i^2$
)

$Students(n)$ can be expressed as an algebraic combination of $Normal(0,1)$ and $Chisquare(n)$

$z \leftarrow Normal(0,1)$
$c \leftarrow Chisquare(n)$
$x \leftarrow z/\sqrt{\frac{c}{n}}$

Verifying Variate Correctness

1. If the sample mean $\bar{x}$ and variance s^2 (or standard deviation) are **inconsistent** with μ and σ^2, then doubt the variate algorithm.
   Not a sufficient test for correctness.
2. Comparing the theoretic pdf $f(x)$ with a small bin size histogram of variate X_is is **sufficient**, but care must be taken around quantization error.
3. Comparing theoretic cdf $F(x)$ with variate cdf is **sufficient**.
   Statistical rules are well known from the literature.

CRV Truncation

From $\mathcal{X} = \{s, s+1, \ldots, a, \ldots, b, b+1, \ldots, t\}$ **to** $\mathcal{X}_t = \{a, \ldots, b\}$

Used for all the same reasons as truncation with DRVs.

Constrained inversion is the preferred technique, in which case the tail probabilities are

$$\alpha = F(a) \qquad \beta = 1 - F(b)$$

and $u \leftarrow Uniform(F(a), F(b))$.

CDF modification can be used as well, in which case

$$f_t(x) = \underbrace{\frac{f(x)}{F(b) - F(a)}}_{\text{normalizes area to 1}} \qquad F_t(x) = \frac{F(x) - F(a)}{F(b) - F(a)} \quad x \in \mathcal{X}_t$$

$$\left(\text{because } F_t(x) = \int_a^x f_t(w)\,\mathrm{d}w = \int_a^x \frac{f(w)}{F(b) - F(a)}\,\mathrm{d}w = \frac{1}{F(b) - F(a)} \int_a^x f(w)\,\mathrm{d}w = \frac{F(x) - F(a)}{F(b) - F(a)}\right)$$