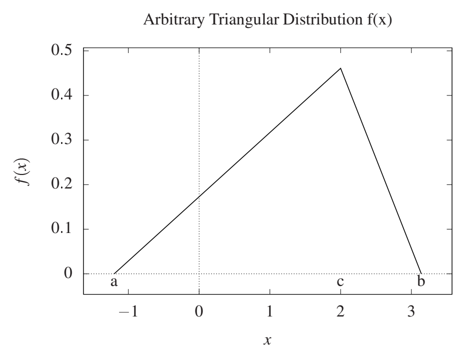
All students should **review** §7.1 and **read** § 7.2–7.3.2 before beginning these problems and in preparation for the next lecture.

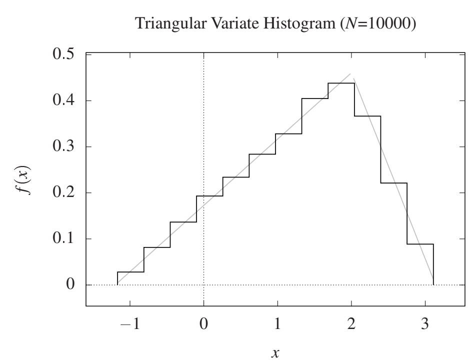
The following numbered questions should be split across your group and the solutions discussed during the next lecture period. Students should review the learning goals for the day, determine which are applicable to their questions and provide answers or commentary to their group members. When using the Internet to formulate answers (some questions may require this), keep track of **where** you find your information on the web. You may be asked for, and are expected to have (in Email-able form), URLs supporting your investigations.

1. Provide an empirical example that **reservoir sampling (algorithm 6.5.6)** works. Let your population be upper case letters (`A`–`Z`, lower case letters (`a`–`z`), and digits (`0`–`9`). Sample 10 values from this population 10 000 times (using the reservoir sampling algorithm of course) and generate a discrete data histogram of the probability each population element was selected in a sample.
2. Question 7.1.5 (§7.1.3) Hint: see Example 7.1.3. Also find X's idf by inverting $F(x) = u$. Your result could be used in code to generate many xs whose histogram would closely match X's pdf.
3. Consider the *triangular random variable* presented in the text (figure 7.3.4) and lecture.



   (a) Derive $F(x)$ and the idf $F^{-1}(u)$, confirm your answers with the text.

   (b) Using a Monte Carlo simulation with 10 000 samples, demonstrate empirical evidence that the histogram of your variate values matches the triangular shape of the pdf. Actually plot the pdf over your histogram, it should look something like the figure to the right.



4. Consider the the three variate algorithms given in the text (§7.2.2) for generating $N(0,1)$ distributed values:
   - approximate inversion using $\Phi_a^{-1}(u)$,
   - summation using $Z = \left(\sum_{i=1}^{i=12} Uniform(0,1)\right) - 6$, and
   - the Box-Muller dual trigonometric equations

   For which of these are distribution **truncation by constrained inversion** feasible? Explain why, and (if need be) generate empirical evidence of your claim. **Note:** I'm not expecting students to look up the equations for $\Phi_a^{-1}(u)$ and implement the variate — given the book description of appoximate inversion and knowledge of how constrained inversion works, you should be able to reason out and explain to your group whether the method is feasible.

5. Demonstrate for your group **constrained inversion** of $Exponential(4.0)$ to the supporting set (pdf domain) $\mathcal{X}_t = (3,6)$. Do this using **two** different implementations of the $Exponential(\mu)$ random variate we have used in the course:

$$F_1^{-1}(u) = -\mu \log(1-u) \qquad \text{and} \qquad F_2^{-1}(u) = -\mu \log u$$

Notice that $u \leftarrow Random()$ is "just as random as" $1-u$, so a coder might expect both to perform equally well under constrained inversion.

Do they? Provide convincing evidence (in the way of histograms or empirical CDFs) to your group when you reconvene next lecture. If you conclude the two methods are different, which is correct? **Why?**

Hints: You can find the CDF ($F(x)$) of the exponential distribution in §7.4.2 of the text. Also, you will want to compare your histograms **not to the original** $f(x)$, but to $\frac{f(x)}{F(b)-F(a)}$, since you have artificially inflated the curve in $\mathcal{X}_t$ to make it a proper distribution with area 1.