Chapter 5 teaching notes

Dean P Foster

September 27, 2010

Administrivia

- No class monday

- I’ll hold a review of homework only. No exam questions allowed. (Fairness issues and to discourage people from coming.) So it will probably be short.

Story time

- Paradox: Two numbers between -one and one written in two envelops.

- Nice strategy, go with the bigger interval. Works more than 1/2 the time for all distributions. Equivalent to, “pick in the direction of 1/2.”

- Can be defeated.

- Fix the defeat by randomizing the 1/2 point. Always works!
• Now, do it for a whole real line.

• Want a compress the whole line to -1,1. Nice trick, $\tan^{-1}()$

• Creates Cauchy.

• It has high probability of generating large numbers.

**Simulation tricks**

Random variables generated by bernulli trials:

• Can just simulate the bernulli trials:
  
  – Do $n$ trials and count success for binomial
  – Count tosses until first success for geometric
  – Count tosses until kth success for negative binomial

• But for $p \approx 0$ this can be slow
  
  – Say $p = 1/million$, how long until first success?
  – About a million tosses

• Trick: use a continuous distribution to approximate the answer

• For binomial, use a normal

• For geometric, use a exponential

• For negative binomial, use a sum of “fast geometrics”
**Exponential distribution**

- Limit of geometric distribution
- Generated by $-\frac{1}{\lambda} \log(1 - U)$.
- Where did this come from?
  - Theorem: For $X$ with continuous CDF $F$, the $F(X)$ is a uniform.
  - Proof: $P(U < u) = P(U < F(x)) = P(F(X) < F(x)) = P(X < x) = F(X) = u$.
- The CDF for an exponential is $F(x) = 1 - e^{-\lambda x}$
- $x = -1/\lambda \log(1 - F(x))$

see [wiki on inverse sampling](https://en.wikipedia.org/wiki/Inverse_sampling)

**Cool properties of exponential**

- Model for waiting times
- How long do you wait until next person arrives in a line? Exponential.
- Related questions:
  - How long until 10 more people arrive? (Sum of 10 exponentials, called a gamma distribution.)
  - How may arrive in 20 minutes? (How many exponentials are needed to generate at least a 20 minute total wait? Called Poisson.)
Normal

- Most important distribution in probability
- Good model for lost of stuff
- approximately normal examples
  - Binomial
  - negative binomial
  - Poisson
  - Gamma
  - Everything but exponential! (Not really, also Cauchy)
- It would be a good world if it were simple, it isn’t.

How to simulate a normal?

- Use an exponential!
- Flip coin to generate sign.
- Draw an exponential.
- Subsample by ratio of normal to exponential
- Some math can make this as tight as possible
- Wiki on rejectin sampling
Some math:

\[-x^2/2 \leq -|x| + 1/2 \quad \text{(Draw the picture)}\]
\[-x^2/2 \leq -x + 1/2 \quad \text{(for positive } x\text{)}\]
\[0 \leq (x - 1)^2/2\]

\[f_{\exp}(x) = e^{-|x|/2}\]
\[f_{\text{norm}}(x) = ke^{-x^2/2}\]
\[k \frac{f_{\text{norm}}(x)}{f_{\exp}(x)} = e^{-(|x|-1)^2/2}\]

\[\{U < k \frac{f_{\text{norm}}(x)}{f_{\exp}(x)}\} = \{- \log(U) > (|x| - 1)^2/2\}\]

Conclusions

- Read the chapter, do the homework.
- There are whole books about special distributions
- The ones we discussed are useful in all fields
- But there are many specialized to particular fields:
  - Physics has Boltzmann, Maxwell and Rayleigh distributions
  - Finance has a log-normal
  - Statistics chi-squared (and many others)
  - Survival analysis (when will you die) has hazard models
Figure 1: It is easy to sample from an exponential distribution (use $-\ln(U)$). If we attach a random sign, we get a Laplace distribution shown in yellow. If we then subsample with probability $e^{-|x|-1)^2}$ we will get the distribution shown in green—which is a normal distribution. Note: The relative area of the two distributions is $\sqrt{\pi/2e}$.

In psudo code this looks like

\[
\begin{align*}
\text{repeat} \\
X &= -\log(rand()) \\
\text{if}(rand() < .5) \quad X &= -X \\
\text{until} \quad -2 * \log(rand()) < (|X| - 1.0)^2
\end{align*}
\]

Where $rand()$ is a uniform random number generate.