CS 286a | **HW 3: Practicing with Heavy-tails** | **Guru: Adam Wierman**

Assigned: 2/17/09 | Due: 2/25/09 noon

I encourage you to discuss these problems with others, but you need to write up the actual solutions alone. At the top of your homework sheet, please list all the people with whom you discussed. Crediting help from other classmates will not take away any credit from you. Also, please limit your use of the web as it is difficult to develop good problems, and so enough search will probably lead you to find the solutions online.

1 Prep for student presentations [0 points]

I expect everyone to read at least the introduction of the papers being presented before the presentations.

2 Important properties of heavy-tailed distributions [30 points]

In this problem, the goal is to contrast heavy-tailed and light-tailed distributions with respect to two important properties that we didn't discuss too much in class. For each part, contrast (i) an exponential, (ii) a Normal with $\sigma^2 = 1$, (iii) a Weibull with $\alpha = 0.5$, and (iv) a Pareto with $\alpha = 1.2$. In each case, set $E[X] = 1$ for consistency.

(a) One way to interpret the central limit theorem is that it tells us that deviations from the mean are of size $\sqrt{n}$. That is, for i.i.d. r.v.s X_i,

$$S_n := \sum_{i=1}^{n} X_n \approx nE[X] + O(\sqrt{n}).$$

In this problem, I'd like you to compare the scale of deviations between heavy-tailed and light-tailed distributions.

In particular, make a plot showing n on the x-axis and $S_n - nE[X]$ on the y-axis for each of the distributions (i)-(iv). Comment on the results, e.g., discuss the order of magnitude of the deviation under each of the distributions. How does this relate to what we discussed in class?

(b) Another important distinguishing feature between heavy-tailed and light-tailed distributions is the "weight" of the tail. By this I mean the percentage of the load of the distribution that is in the largest 1% of the jobs. (Think of the Pareto principle.) To get a feel for this, make a plot of $\bar{F}(x)$ vs. $\int_x^\infty tf(t)dt/E[X]$ for each of the distributions (i)-(iv). Include a marker for the case of $\bar{F}(x) = .01$.

3 Practicing with the terminology [35 points]

(a) Prove or disprove that the independent sum of two light-tailed random variables is light-tailed. (Light-tailed distributions are those that aren't heavy-tailed.)

(b) Show that if $L_1(x)$ and $L_2(x)$ are both slowly varying then $L_1(x) + L_2(x)$ is slowly varying.

(c) Show that the Pareto distribution is subexponential.

(d) Derive $Pr(\max(X_1, \ldots, X_n) > t)/Pr(X_1 + \ldots + X_n > t)$ under the exponential distribution. Relate this to the definition of subexponential distributions.

(e) Prove that $X \sim RV(\alpha)$ implies that $E[X^n] = \infty$ for $n > \alpha$ and $E[X^n] < \infty$ for $n < \alpha$.

4 Getting ready for large deviations [35 points]

In this problem we'll derive Chernoff bounds, which are the basic tool of large deviations analysis that we'll be learning in a few lectures. Many of you claim to have seen Chernoff bounds before – in that case, this problem will be easy! But, this will likely still provide a new view of it... And, for those of you who haven't seen Chernoff bounds before, this should be a nice introduction to them.

The idea of large deviations analysis is to study the behavior of $Pr(X > t)$ for large t. It's clear why we need to do this in the case of scheduling and queueing. We can't derive $Pr(X > t)$ exactly, so instead we look at an asymptotic version of it.

One simple tool for accomplishing this is Chernoff bounds.

(a) Derive the general Chernoff bound using Markov's inequality. In particular, prove that

$$\log Pr(X > t) \leq \min_{s>0}\{(\log E[e^{sX}]) - st\}.$$

Hint: $Pr(X > t) = Pr(e^{sX} > e^{st})$.

The most important difference between this "Chernoff bound" and Markov's inequality is that the Chernoff bound is in terms of the moment generating function, which encodes the entire probability distribution. In contrast, Markov's inequality uses just the mean of the distribution.

(b) The general form we have derived above is now a useful tool to apply to any distribution we feel like. Derive the Chernoff bound for an Exponential random variable. Discuss the tightness of the bound – one way to do this is using a plot of the bound and the true value.

(c) We typically apply Chernoff bounds to the i.i.d. sum of random variables, $S_n = X_1 + \ldots + X_n$. Prove that

$$\log Pr(S_n > nt) \leq -n \max_{s>0}\{st - \log E[e^{sX}]\}$$

Typically,

$$I(x) := \max_{s>0}\{st - \log E[e^{sX}]\}$$

is referred to as the "decay rate." *Explain why.*

Additionally, $\log E[e^{sX}]$ is referred to as the "cumulant generating function." This is because it can be used to generate the "cumulant moments" just like the moment generating function can be used to generate the raw moments. Cumulants are extremely interesting things, and I recommend looking up more about them on wikipedia.

(d) Use the above to derive a Chernoff bound for $S_n \sim Binomial(n, p)$. Discuss the tightness of the bound. Discuss the tightness of the Chernoff bound – one way to do this is using a plot of the bound and the true value.

(e) For the case of $S_n \sim Binomial(n, p)$, derive a bound using Markov's inequality and Chebychev's Inequality. Contrast these bounds with the Chernoff bound.

(f) I commented above that Chernoff bounds are useful for large deviations analysis, but they seem to provide bounds on $Pr(X > t)$ for all t, not just $t \to \infty$. Why might they be most useful as $t \to \infty$? Be specific and refer to the earlier parts of the problem if possible.

(g) When might it be impossible to use Chernoff bounds?