Homework Assignment 2

Due in class, Wednesday Feb 21st

SDS 384-11 Theoretical Statistics

1. Show that Markov's inequality is tight.

   (a) Give an example of a non-negative random variable X and a value $k > 1$ such that $P(X \geq kE[X]) = 1/k$.
   Take

   $$X = \begin{cases} 0 & \text{w.p. } 1 - 1/k \\ k & \text{w.p. } 1/k \end{cases}$$

   $E[X] = 1$ and $P(X \geq k) = 1/k$.

   (b) Give an example of a random variable X (with $E[X] > 0$) and a value $k > 1$ such that $P[X \geq kE[X]] > 1/k$.

   $$X = \begin{cases} -k & \text{w.p. } 1/k \\ 0 & \text{w.p. } 1 - 3/k \\ k & \text{w.p. } 2/k \end{cases}$$

   $E[X] = 1$ and $P(X \geq k) = 2/k$.

2. Consider a r.v. X such that for all $\lambda \in \Re$

   $$E[e^{\lambda X}] \leq e^{\frac{\lambda^2\sigma^2}{2} + \lambda\mu} \tag{1}$$

   Prove that:

   (a) $E[X] = \mu$. Let $f(\lambda) = E[e^{\lambda X}]$ and let $g(\lambda) = e^{\lambda^2\sigma^2/2 + \lambda\mu}$. We have $f(0) = g(0)$.

   $$f'(0) = \lim_{h \to 0} \frac{f(h) - f(0)}{h} \leq \lim_{h \to 0} \frac{g(h) - g(0)}{h} = g'(0)$$

   But we also have:

   $$f'(0) = \lim_{h \to 0} \frac{f(0) - f(-h)}{h} \geq \lim_{h \to 0} \frac{g(0) - g(-h)}{h} = g'(0)$$

   So $f'(0) = g'(0)$. So we have $E[X] = \mu$.

(b) $\text{var}(X) \le \sigma^2$. First note that for subgaussian R.V's, we have the following moment bound on the higher moments. Take $E[X] = 0$ WLOG. First note that we have:

$$P(|X| > t) \le 2\exp(-t^2/2\sigma^2)$$

$$\begin{aligned} E[|X|^k] &\le \int_0^\infty P(|X| \ge t^{1/k})dt \le 2\int_0^\infty e^{-\frac{t^{2/k}}{2\sigma^2}}dt \\ &= (2\sigma^2)^{k/2}k\int_0^\infty e^{-u}u^{k/2-1}du = (2\sigma^2)^{k/2}k\Gamma(k/2) \le (C\sigma\sqrt{k})^k \end{aligned}$$

Now using the above and Stirling's approximation we have: $f(\lambda) = 1+\lambda^2\text{var}(X)/2+\sum_{k>2}\lambda^k E[X^k]/k! = 1+\lambda^2\text{var}(X)/2+o(\lambda^2)$. So we have for $\lambda \to 0$:

$$1+\lambda^2\text{var}(X)/2 \le 1+\lambda^2\sigma^2+o(\lambda^2)$$

Subtracting 1 from both sides and dividing both sides by λ^2, and then taking $\lambda \to 0$ shows that $\text{var}(X) \le \sigma^2$.

(c) If the smallest value of σ satisfying the above equation is chosen, is it true that $\text{var}(X) = \sigma^2$? Prove or give a counter example. Take $X \sim \mathit{Bernoulli}(p)$. So $E[e^{t(X-p)}] = e^{-tp}(e^t p + 1 - p)$. We know that X is subgaussian. So $\exists\sigma > 0$ s.t. $E[e^{t(X-p)}] \le e^{t^2\sigma^2/2}$. Take $t = 1$. The smallest σ that satisfies the upper bound is $\sigma^2 = 2(-p + \log(pe + 1 - p))$, which is smaller than $p(1-p)$ for $p = .1$.

3. Remember Hoeffding's Lemma? We proved it with a weaker constant in class using a symmetrization type argument. Now we will prove the original version. Let X be a bounded r.v. in $[a, b]$ such that $E[X] = \mu$. Let $f(\lambda) = \log E[e^{\lambda(X-\mu)}]$. Show that $f''(\lambda) \le (b-a)^2/4$. Now use the fundamental theorem of calculus to write $f(\lambda)$ in terms of $f''(\lambda)$ and finish the argument. Take $\mu = 0$ WLOG. Note that $f'(\lambda) = \frac{E[Xe^{\lambda X}]}{E[e^{\lambda X}]}$ and furthermore, $f''(\lambda) = \frac{E[X^2e^{\lambda X}]}{E[e^{\lambda X}]} - \left(\frac{E[Xe^{\lambda X}]}{E[e^{\lambda X}]}\right)^2$. So $f''(\lambda)$ can be thought of as the variance of $X \sim g$ where $g(x) = e^{\lambda x}/E[e^{\lambda x}]p(x)$ where $p(x)$ is the original density of X. Since $p(x)$ has support $[a, b]$, one can easily check that the support of $g(x)$ is also $[a, b]$. So, $f''(\lambda) = \text{var}(X) \le (b-a)^2/4$. Now the fundamental theorem of calculus gives:

$$f(\lambda) = \int_0^\lambda \int_0^t f''(\rho)d\rho dt \le \frac{\lambda^2(b-a)^2}{8}$$

Now if $\mu \ne 0$, then $g(x)$ will have support on $[a-\mu, b-\mu]$, and the rest of the argument goes through almost identically.

4. Bernstein's inequality for bounded i.i.d sequences of random variables $\{X_i\}$ with $|X_i| \le M$ gives: $P(|\sum_i(X_i - E[X_i])| \ge t) \le 2\exp\left(\frac{-t^2/2}{\sum_i \text{var}(X_i)+Mt/3}\right)$. Consider n i.i.d. $X_i \sim \text{Bernoulli}(p_n)$ r.v's. We will consider two cases to study concentration of $\bar{X}_n$ around p_n.

(a) (Dense case) Let $np_n/\log n \to \infty$. Can you apply Hoeffding's bound and Bernstein's inequality to establish concentration of $\bar{X}_n$, i.e. $P(\bar{X}_n \in [p_n(1-\epsilon_n), p_n(1+\epsilon_n)]) = 1-O(1/n)$, where $\epsilon_n \to 0$? Do you prefer one bound over another? Why? Hoeffding's inequality gives:

$$P(|X - E[X]| \geq t) \leq 2e^{-2t^2/n}$$

In the dense case, we take $t = \theta(\sqrt{n\log n})$, we have $t/np = \sqrt{\log n/np^2}$ which only goes to zero when $np \gg \sqrt{n\log n}$. Hoeffding does not work when $\log n \ll np \ll \sqrt{n\log n}$. $P(|\sum_i (X_i - E[X_i])| \geq t) \leq 2\exp\left(\frac{-t^2/2}{np(1-p)+t/3}\right)$ Taking $t = \Theta(\sqrt{np\log n})$, gives $t = o(np)$ and the error probability is also $O(1/n)$. **What happens with Chernoff?**

(b) (Sparse case) Repeat your argument for the case $np_n = c\log n$ where c is some constant not depending on n. Hoeffding will not work here. If you take $t = \sqrt{np\log^{1/2} n}$, the error probability is $exp(-\log^{1/2} n)$ which is not $O(1/n)$ but is $o(1)$, so there is consistency, but with a much larger error probability.