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

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

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) $E[X] = \mu$.

   (b) $\text{var}(X) \leq \sigma^2$.

   (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.

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 $f''(\lambda) \leq (b-a)^2/4$. Now use the fundamental theorem of calculus to write $f(\lambda)$ in terms of $f''(\lambda)$ and finish the argument.

4. Bernstein's inequality for bounded i.i.d sequences of random variables $\{X_i\}$ with $|X_i| \leq M$ gives: $P(|\sum_i (X_i - E[X_i])| \geq t) \leq 2\exp\left(\frac{-t^2/2}{\sum_i \text{var}(X_i) + Mt/3}\right)$. 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 $p_2 n$.

   (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. $P(\bar{X}_n \in [p_n(1-\epsilon_n), p_n(1+\epsilon_n)]) = O(1/n)$, where $\epsilon_n \to 0$? Do you prefer one bound over another? Why?

   (b) (Sparse case) Repeat your argument for the case $np_n = c \log n$ where c is some constant not depending on n.