Homework Assignment 3

Due in class, Wednesday March 7th

SDS 384-11 Theoretical Statistics

1. Suppose that X_1 and X_2 are zero-mean and sub-Gaussian with parameters σ_1 and σ_2 respectively. **Assume that the variance parameters are equal to the sub-gaussian parameters, i.e. $E[X_1^2] = \sigma_1^2$ and $E[X_2^2] = \sigma_2^2$. This is needed for part (a) and (c) uses part (a).**

   (a) Show that the MGF of $V := X_1^2 - E[X_1^2]$ can be bounded as $E[e^{tV}] \leq e^{2\sigma_1^4 t^2}$ for $0 \leq t \leq 1/4\sigma_1^2$. *Hint: write the mgf in terms of X_1 and an independent standard normal.*

   $$E[e^{t(X_1^2)}] = E_{X_1,Z}[e^{\sqrt{2t}X_1 Z}] \leq E_Z[e^{t\sigma_1^2 Z^2}]$$
   $$E[e^{t(X_1^2-\sigma_1^2)}] \leq E_Z[e^{t\sigma_1^2(Z^2-1)}]$$

   Since $Z^2 - 1$ is subexponential $(2, 4)$ the result follows.

   (b) If X_1 and X_2 are not independent, show that $X_1 + X_2$ is sub-Gaussian with parameter at most $\sqrt{2(\sigma_1^2 + \sigma_2^2)}$.

   $$E[e^{t(X_1+X_2)}] \overset{Cauchy-Schwarz}{\leq} \sqrt{E[e^{2tX_1}]E[e^{2tX_2}]}$$
   $$\leq e^{t^2(\sigma_1^2+\sigma_2^2)} = e^{t^2\sigma^2/2}$$

   where $\sigma = \sqrt{2(\sigma_1^2 + \sigma_2^2)}$.

   (c) If X_1 and X_2 are independent, show that X_1X_2 is sub-exponential with parameters $(\sqrt{2}\sigma_1\sigma_2, \sqrt{2}\sigma_1\sigma_2)$. **It seems that there is a typo in Martin's book, which is fixed. Thanks to Mohamed.** Let V be defined as in part (a).

   $$E[e^{tX_1X_2}] \leq E[e^{t^2\sigma_2^2 X_1^2/2}] = E[e^{t^2\sigma_2^2 V/2}]e^{t^2\sigma_1^2\sigma_2^2/2}$$
   $$\leq e^{\sigma_1^4\sigma_2^4 t^4/2 + t^2\sigma_1^2\sigma_2^2/2} \qquad \text{For } t^2 < \frac{1}{2\sigma_1^2\sigma_2^2}$$
   $$\leq e^{t^2(2\sigma_1^2\sigma_2^2)/2}$$

2. Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples of random variable with density f on the real line. A standard estimate of f is the kernel density estimate

   $$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where $K : \Re \to [0, \infty)$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(t)dt = 1$, and h is a bandwidth parameter. We will measure the quality of $\hat{f}$ using

$$\|\hat{f} - f\|_1 := \int_{-\infty}^{\infty} |\hat{f}(t) - f(t)|dt.$$

Prove that:

$$P(\|\hat{f} - f\|_1 \geq E\|\hat{f} - f\|_1 + \delta) \leq e^{-cn\delta^2},$$

where c is some constant.

This seems like something suited for McDiarmid or the bounded differences inequality. So we will first calculate how big the differences are.

$$\begin{aligned}
&|\|\hat{f}_{X_1,\ldots,X_n}(t) - f(t)\|_1 - \|\hat{f}_{X_1',\ldots,X_n}(t) - f(t)\|_1| \\
&\leq \int |\hat{f}_{X_1,\ldots,X_n}(t) - \hat{f}_{X_1',\ldots,X_n}(t)|dt \\
&\leq 1/nh \int \left| K\left(\frac{x - X_i}{h}\right) - K\left(\frac{x - X_i'}{h}\right)\right| dt \leq 2/n
\end{aligned}$$

So, now we can use McDiarmid's inequality to get the above result.

3. Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of Bernoulli variables with parameter $\alpha \in (0, 1/2]$, and consider the binomial random variable $Z_n = \sum_i X_i$. We want to prove for any $\delta \in (0, \alpha)$,

$$P(Z_n \leq \delta n) \leq \exp(-nKL(\delta||\alpha)) \qquad KL(\delta||\alpha) := \delta \log \frac{\delta}{\alpha} + (1 - \delta) \log \frac{1 - \delta}{1 - \alpha}$$

where $KL(p, q)$ is the Kullback-Leibler divergence between two bernoullis with parameters p, q respectively. Show that the above is strictly better than Hoeffding's inequality.

$$P(Z_n \leq \delta n) \leq \inf_{\lambda \leq 0} E[e^{\lambda Z_n - \lambda\delta n}] = \inf_{\lambda \leq 0} e^{-\lambda\delta n}(e^{\lambda}\alpha + (1 - \alpha))^n$$

Typically we do some approximations at this point. However, we can directly minimize the above function w.r.t λ. This gives

$$\begin{aligned}
&\frac{\partial}{\partial\lambda}\left(-\lambda\delta n + n\log(e^{\lambda}\alpha + (1 - \alpha))\right) = 0 \\
&\lambda = \log \frac{\delta(1 - \alpha)}{(1 - \delta)\alpha}
\end{aligned}$$

Note that this is also smaller than zero since $\alpha > \delta$ and $1 - \alpha < 1 - \delta$. Plugging this in, we have:

$$P(Z_n \leq \delta n) \leq e^{-\lambda\delta n}(e^{\lambda}\alpha + (1 - \alpha))^n$$

Plugging in the optimal λ gives the result.

4. Now we will prove a lower bound on the binomial tail to show that indeed what you derived in the last question is sharp upto polynomial factors. Define $m = \lfloor n\delta \rfloor$ and $\delta' = \frac{m}{n}$.

   (a) Prove $\frac{1}{n}\log P(Z_n \leq \delta n) \geq \frac{1}{n}\log\binom{n}{m} + \delta' \log\alpha + (1-\delta')\log(1-\alpha)$.
   $P(Z_n \leq \delta n) \geq P(Z_n \leq m) \geq \binom{n}{m}\alpha^m(1-\alpha)^{n-m}$. Taking a log and dividing both sides by n gives the answer.

   (b) Show that

   $$\frac{1}{n}\log\binom{n}{m} \geq -\delta'\log\delta' - (1-\delta')\log(1-\delta') - \frac{\log(n+1)}{n}$$

   Hint: Use the fact that for $Y \sim Bin(n, m/n)$ $P(Y = k)$ is maximized at $k = m$.
   We will use the fact that the largest value of a binomial$(n, m/n)$ PMF has to be larger than a Uniform distribution on $\{0, n\}$. Because if it was smaller, then the binomial PMF will sum to something smaller than one, i.e.,

   $$\binom{n}{m}(\delta')^m(1-\delta')^{n-m} \geq \frac{1}{n+1}.$$

   Taking a logarithm and dividing by n gives the result.

   (c) Now show that

   $$P(Z_n \leq \delta n) \geq \frac{1}{n+1}\exp(-nKL(\delta'||\alpha))$$

   Note that the original question had δ here. Asymptotically this is not incorrect, since δ' and δ are asymptotically the same. But just to avoid confusion, I am replacing this with δ'. Thanks to Jinxie for pointing it out.

   Plugging in part (b)'s lower bound into part (a) and exponentiating both sides immediately gives the final result.