36-788, Fall 2015
Homework 1

Due Sep 17.

1. (Mill's ratio). Let $\Phi\colon \mathbb{R} \to [0,1]$ the c.d.f. of the standard Gaussian distribution on $\mathbb{R}$ and ϕ its p.d.f..

   (a) Prove that, for all $x > 0$,
   $$\frac{x}{1+x^2}\phi(x) \leq \Phi(x) \leq \frac{1}{x}\phi(x)$$

   (b) Prove that, for all $x > 0$,
   $$\Phi(x) \leq \frac{1}{2}\exp\left(-x^2/2\right).$$

2. Let $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ be a random vector with covariance matrix Σ such that $\frac{X_i}{\sqrt{\Sigma_{i,i}}}$ is sub-Gaussian with parameter ν^2, for all $i = 1, \ldots, d$. Assume we observe n i.i.d. copies of X and compute the empirical covariance matrix $\widehat{\Sigma}$. Show that, for all $i, j \in \{1, \ldots, d\}$,
   $$\mathbb{P}\left(\left|\widehat{S}_{i,j} - \Sigma_{i,j}\right| > \epsilon\right) \leq C_1 e^{-\epsilon^2 n C_2},$$
   for some constants C_1 and C_2. Conclude that
   $$\max_{i,j}\left|\widehat{S}_{i,j} - \Sigma_{i,j}\right| = O_P\left(\sqrt{\frac{\log d}{n}}.\right)$$
   You may want to consult these references:

   - Lemma 12 in Yuan. M. (2010). High Dimensional Inverse Covariance Matrix Estimation via Linear Programming, JMLR, 11, 2261-2286.
   - Lemma 1 in Ravikumar, P., Wainwright, M.J., Raskutti, G. and Yu, B. (2011). EJS, 5, 935-980.
   - Lemma A.3 in Bickel, P.J. and Levina, E. (2008). Regularized estimation of large covariance matrices, teh Annals of Statistics, 36(1), 199-227.

3. (Sampling with replacement). Let $\mathcal{X}$ a finite set with N elements. Let $X_1, \ldots, X_n$ be a random sample without replacement from $\mathcal{X}$ and $Y_1, \ldots, Y_n$ be a random sample with replacement from $\mathcal{X}$. Show that, for any convex function $f\colon \mathbb{R}^n \to \mathbb{R}$,
   $$\mathbb{E}\left[f\left(\sum_{i=1}^n X_i\right)\right] \leq \mathbb{E}\left[f\left(\sum_{i=1}^n Y_i\right)\right].$$
   Use this result to show that all the inequalities derived for the sums of independent random variables $\{Y_1, \ldots, Y_n\}$ using Chernoff's bounding techniques remain true also for the sums of the X_i's. (see Hoeffding, W. (1963). Probility Inequalities for sums of Bounded Random Variables, by W. Hoeffding, *JASA*, 58, 13–30., 1963).

4. (Moments versus Cernoff bounds). Show that moment bounds for tail probabilities are always better than CrameérChernoff bounds. More precisely, let Y be a nonnegative random variable and let $t > 0$.

The best moment bound for the tail probability $\mathbb{P}Y \geq t$ is $\min_q \mathbb{E}[Y^q]t^q$ where the minimum is taken over all positive integers. The best CramérChernoff bound is $\inf_{\lambda>0} \mathbb{E}e^{\lambda(Yt)}$. Prove that

$$\min_q \mathbb{E}[Y^q]t^q \leq \inf_{\lambda>0} \mathbb{E}e^{\lambda(Yt)}.$$

(See Philips, T.K. and Nelson, R. (1995). The moment bound is tighter than Chernoffs bound for positive tail probabilities. *The American Statistician*, 49, 175-178.)

5. (Concentration function of the standard Gaussian). Here is a way to get the (dimension free) concentration function for the standard normal in $\mathbb{R}^n$, with suboptimal constants. Let γ_n denote the standard Gaussian distribution in $\mathbb{R}^n$ and $A \subset \mathbb{R}^n$ be measurable. Prove that

$$\int_{\mathbb{R}^n} e^{d(x,A)} d\gamma_n(x) \leq \frac{1}{\gamma_n(A)},$$

which implies that, for any A with $\gamma_n(A) \geq 1/2$,

$$\gamma_n(A_r^c) \leq 2e^{r^2/4}.$$

Above, $d(x, A) = \inf_{y \in A} \|x - y\|$ and $A_r = \{x \colon d(X, A) < r\}$, $r > 0$.

Proceed in this way: use the following theorem with $f(x) = e^{d(x,A)^2/4}\gamma_n(x)$, $g(x) = 1_{x \in A}\gamma_n(x)$ and $h(x) = \gamma_n$.

Theorem 0.1 (Prépoka-Leinder Inequality) *If f and g and h are non-negative measurable functions on $\mathbb{R}^n$ and $\lambda \in (0, 1)$ is such that, for all x and y in $\mathbb{R}^n$,*

$$h\left((1-\lambda)x + \lambda y\right) \geq f(x)^{1-\lambda} g(y)^{\lambda},$$

then,

$$\int_{\mathbb{R}^n} h(x)dx = \left(\int_{\mathbb{R}^n} f(x)dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g(x)dx\right)^{\lambda}.$$

See Theorem 8.1 in: K. Ball, An elementary introduction to modern convex geometry, in: Flavors of Geometry, pp. 158, Math. Sci. Res. Inst. Publ. Vol. 31, Cambridge Univ. Press, Cambridge, 1997.

6. Bounds bounds bounds...

   (a) Suppose that, for all $\epsilon > 0$,

   $$\mathbb{P}\left(|X| \geq \epsilon\right) \leq c_1 e^{-c_1 n \epsilon^a},$$

   where $a \in \{1, 2\}$. Show that

   $$\mathbb{E}\left[|X|\right] \leq c_3 n^{-1/a}$$

   and express c_3 as a function of c_1 and c_2.

   (b) (From Bernstein exponential inequality to high probability bounds). Suppose that, for all $\epsilon > 0$, and some positive constants a, b, c and d,

   $$\mathbb{P}\left(|X| \geq t\right) \leq a\exp\left\{-\frac{nb\epsilon^2}{c + d\epsilon}\right\}.$$

   Then show that, for any $\delta \in (0, 1)$,

   $$|X| \leq \sqrt{\frac{c}{nb} \ln \frac{a}{\delta}} + \frac{d}{nb} \ln \frac{a}{\delta},$$

   with probability at least $1 - \delta$.

7. Let X be distributed like a $N_n(0, I_n)$, where I_n is the n-dimensional identity matrix. Show that, for any $\epsilon \in (0,1)$

$$\mathbb{P}\left(\left|\|X\|^2 - n\right| \geq \epsilon\right) \leq e^{-n\epsilon^2/8}.$$

This results says that, in high dimensions, X is tightly concentrated around a sphere of radius $\sqrt{n}$. This may seem counterintuitive, since the density of X is maximal at 0 and decays exponentially fast as we move away from the origin. So what's going on? (Hint: think about the fact that most of the volume of the ball in $\mathbb{R}^n$ of radius $\sqrt{n}$ concentrates around the sphere of radius $\sqrt{n}$ enclosing it...).

See, e.g., Lemma 2 in A Probabilistic Analysis of EM for Mixtures of Separated, Spherical Gaussians, by S. Dasgupta and L. Schulman, JMLR, 8, 203–26, 2007.