E2 207 – Concentration Inequalities, Aug-Dec 2017
Homework #2

---

(1) *Distinct values in an iid sequence*
Let $X_1, \ldots, X_n$ be iid random variables taking positive integer values, and let Z_n be the number of distinct values taken by these random variables.
(a) Show that $\lim_{n\to\infty} \mathbb{E}[Z_n/n] = 0$, or in other words, $\mathbb{E}[Z_n/n] = o(n)$.
(b) Does Z_n satisfy the bounded differences property? If so, what can you conclude about the variance of Z_n (as a function of n)?
(c) Show that Z_n is in fact a self-bounding function. What can you now conclude about its variance as a function of n? Compare this to the conclusion of the previous part.

(2) *Order Statistics and variance*
Let $X_1, \ldots, X_n$ be a sequence of independent random variables, and $X_{(1)} \leq X_{(1)} \leq \cdots \leq X_{(n)}$ a sorted version of the sequence ($X_{(i)}$ is known as the i-th *order statistic*).
(a) Show that $\mathrm{Var}[X_{(n)}] \leq \mathbb{E}\left[\left(X_{(n)} - X_{(n-1)}\right)^2\right]$.
(b) Compute the LHS and RHS of the inequality above, when all the X_i are iid Exponential(1).
(c) Repeat for all the X_i iid Uniform($[0, 1]$).

(3) *Jackknife estimators*
Let $X_1, \ldots, X_n$ be an iid sequence of random variables (a "sample" of size n in statistics terminology). Suppose one has designed, for any n, an estimator $T_n \equiv T_n(X_1, \ldots, X_n)$ for a scalar parameter $\theta \in \mathbb{R}$ of the common probability distribution of the X_i. One often wants to know the quality of the estimator T_n (using only the sample). The jackknife is a method to estimate the bias $\mathbb{E}[T_n] - \theta$ and variance $\mathbb{E}[(T_n - \mathbb{E}[T_n])^2]$ of the estimator T_n.
For each $i \in [n]$, define the ith *pseudo-value*

$$Y_i := n\, T_n(X_1, \ldots, X_n) - (n-1)\, T_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

The *jackknife estimate of the bias* of T_n is defined to be the difference between T_n and the sample mean of the pseudo-values, i.e., $\hat{B}_n := T_n - \frac{1}{n}\sum_{i=1}^n Y_i$, while the *jackknife estimate of the variance* of T_n is defined to be the sample variance of the pseudo-values Y_i, i.e., $\hat{V}_n := \frac{1}{n-1}\sum_{i=1}^n \left(Y_i - \frac{1}{n}\sum_{j=1}^n Y_j\right)^2$. (The general principle is to imagine the pseudo values Y_i as representing "iid copies of T_n" and compute standard statistics on them.)
(a) Compute the jackknife estimate of the bias of the sample mean estimator $T_n := \frac{1}{n}\sum_{i=1}^n X_i$.
(b) For any estimator T_n, show that $\hat{V}_n$, the jackknife estimator of the variance of T_n, is always positively biased, i.e., $\mathbb{E}\left[\hat{V}_n\right] - \mathrm{Var}[T_n] \geq 0$.

(4) *Gradients of Lipschitz functions*
Show that if a function $f : \mathbb{R}^n \to R$ is Lipschitz, with respect to the standard Euclidean norm on R^n, and differentiable, then the norm of its gradient is bounded by 1.

(5) *Log-Sobolev is stronger than Poincaré*
Let $f : \mathbb{R}^n \to R$ be a bounded, continuously differentiable function. Show that the Gaussian logarithmic Sobolev inequality for $f(X)$, with $X \sim \mathcal{N}(0, I_{n\times n})$, implies the Gaussian Poincaré inequality for $f(X)$.

(6) *Log-Sobolev for the exponential distribution*
Let X be an exponentially distributed random variable with parameter 1, and let $f : [0, \infty) \to \mathbb{R}$ be a continuously differentiable function. Show that $\mathrm{Ent}(f(X)^2) \leq 4\,\mathbb{E}\,[X(f'(X))^2]$.