Statistics 521, Problem Set 1

Wellner; 9/26/12

Reading:	Shorack, PfS, Chapter 1, pages 1 - 17;
Due:	Wednesday, October 3, 2012.

1. (a) Suppose that $\{\mathcal{A}_n\}$ is an increasing sequence of algebras, i.e. $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ for all $n \geq 1$. Show that $\cup_{n=1}^{\infty}\mathcal{A}_n$ is an algebra.
(b) Suppose that the $\mathcal{A}_n$ of (a) are $\sigma-$algebras. Show by constructing a counter-example that $\cup_{n=1}^{\infty}\mathcal{A}_n$ need not be a $\sigma-$algebra.

2. Write out a proof of Proposition 1.1(b), PfS, page 3.

3. PfS, Exercise 1.1.1, PfS, page 4.

4. PfS, Exercise 1.1.2, PfS, page 8.

5. PfS(2000), Exercise 9.1.4, page 182 and PfS(2012), Exercise A.1.4, page 428. Suppose that $X_n \sim \text{Binomial}(n, p_n)$ where $np_n \to \lambda > 0$. Show that
$$P(X_n = k) \to \frac{\lambda^k}{k!}\exp(-\lambda) = P(Y = k)$$
where $Y \sim \text{Poisson}(\lambda)$; this implies that $X_n \to_d Y$.

6. Let $I = P(Z \geq 2) = .02275...$ where $Z \sim N(0,1)$. Thus $I = \int h(x)f(x)dx$ there $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standard normal density and $h(x) = 1_{[x\geq 2]}$. The basic Monte Carlo estimator is $\hat{I}_1 = n^{-1}\sum_{i=1}^n h(X_i)$ where $X_1, \ldots, X_n$ are i.i.d. $N(0,1)$. When I carried this out I found $\hat{I}_1 = .04$ and $\widehat{Var(\hat{I}_1)} = .0239/100$. Note that most observations are "wasted" in that they are not near the right tail. Now we try again with "importance sampling": let g be the $N(3,1)$ density. When we sample from g, the corresponding estimator of I is given by $\hat{I}_2 = n^{-1}\sum_{i=1}^n f(X_i)h(X_i)/g(X_i)$. When I do this I find $\hat{I}_2 = .0239...$ and $\widehat{Var(\hat{I}_2)} = .00218/100$. Note that $\widehat{Var(\hat{I}_1)}/\widehat{Var(\hat{I}_2)} = .0239/.00218 \approx 10.9$.
(a) Show that
$$Var(\hat{I}_1) = p(1-p)/n = .02275(1 - .02275)/100 = 0.0222.../100$$

where $p = I = P(Z \geq 2) = .02275....$
(b) Show that

$$Var(\hat{I}_2) = n^{-1} Var_g(f(X)h(X)/g(X)) = .001805.../100,$$

and hence the variance has been reduced by a factor of 12.3.
Hint: You may compute

$$\int_{-\infty}^{\infty} \left\{ h^2(x) f^2(x)/g(x) \right\} dx$$

numerically in (b).

7. (Optional bonus problem). Suppose that X is a random variable with density f and $h : \mathbb{R} \to \mathbb{R}$ satisfies $E|h(X)| = \int_{\mathbb{R}} |h(x)| f(x) dx < \infty$. Suppose that we want to estimate (or approximate) $I = Eh(X) = \int_{\mathbb{R}} h(x) f(x) dx$ based on sampling $Y_1, \ldots, Y_n$ from a density g on $\mathbb{R}$, and then forming $\hat{I}_n = n^{-1} \sum_{i=1}^{n} h(Y_i) f(Y_i)/g(Y_i)$. Show that the choice $g(y) = |h(y)| f(y) / \int |h(z)| f(z) dz$ minimizes the variance of $\hat{I}_n$.