14.381: Statistics

Problem Set 1

Problems 1-2 and 4-5 are for practice. They will be discussed at recitation. One of the problems from the problem sets will appear on the mid-term exam.

1. Let X and Y be random variables with finite variances.

   (i) Show that
   $$\min_{g(\cdot)} E\left(Y - g(X)\right)^2 = E\left(Y - E\left(Y \mid X\right)\right)^2,$$
   where $g(\cdot)$ ranges over all functions.

   (ii) Assume $m(X) = E(Y|X)$ and write $Y = m(X) + e$. Show that $Var(Y) = Var(m(X)) + Var(e)$.

   (iii) If $E(Y|X = x) = a + bx$ find $E(YX)$ as a function of moments of X.

2. Show that if a sequence of random variables ξ_i converges in distribution to a constant c, then $\xi_i \xrightarrow{p} c$.

3. *(The required problem)* Let $\{X_i\}$ be independent Bernoulli (p). Then $EX_i = p$, $Var(X_i) = p(1-p)$. Let $Y_n = \frac{1}{n}\sum_{i=1}^n X_i$.

   (a) Describe the asymptotic behavior of Y_n.

   (b) Show that for $p \neq \frac{1}{2}$ the estimated variance $Y_n(1 - Y_n)$ has the following limit behavior
   $$\sqrt{n}(Y_n(1 - Y_n) - p(1 - p)) \Rightarrow N(0, (1 - 2p)^2 p(1 - p)).$$

   (c) Prove that if (i) $\frac{\sqrt{n}}{\sigma}(\xi_n - \mu) \Rightarrow N(0, 1)$ (ii) g is twice continuously differentiable: $g'(\mu) = 0$, $g''(\mu) \neq 0$, then
   $$n(g(\xi_n) - g(\mu)) \Rightarrow \sigma^2 \frac{g''(\mu)}{2} \chi_1^2.$$

Note. You may assume that g has more derivatives, if it simplifies your life. Use O_p and o_p notation wherever possible.

Note: χ_1^2 is a chi-square distribution with 1 degree of freedom. Let $\xi_1, \ldots, \xi_p$ be i.i.d. $N(0,1)$, then $\chi_p^2 = \sum_{i=1}^p \xi_i^2$.

(d) Show that for $p = \frac{1}{2}$

$$n\left[Y_n(1-Y_n) - \frac{1}{4}\right] \Rightarrow -\frac{1}{4}\chi_1^2$$

Curious fact: Note that $Y_n(1-Y_n) \leq \frac{1}{4}$, that is, we always underestimate the variance for $p = \frac{1}{2}$.

4. (Multivariate limit theorems) Let $\mathbf{X} = (X_1, ..., X_m)'$ and $\mathbf{X}_n = (X_{n1}, ..., X_{nm})'$ be m-dimensional random vectors. Define a norm $\|\mathbf{X}\| = \sqrt{X_1^2 + ... + X_m^2}$.

   (a) Show that $E\|\mathbf{X}\| < \infty$ if and only if $E|X_i| < \infty$ for all $i = 1, .., m$.

   (b) Define $\mathbf{X}_n \to^p \mathbf{X}$ if for any $\varepsilon > 0, \lim_{n\to\infty} P\{\|\mathbf{X}_n - \mathbf{X}\| > \varepsilon\} = 0$. Show $\mathbf{X}_n \to^p \mathbf{X}$ if and only if $X_{ni} \to^p X_i$ for all $i = 1, ..., m$.

   (c) Define $\mathbf{X}_n \Rightarrow \mathbf{X}$ if and only if for any non-random m-dimensional vector λ such that $\|\lambda\| = 1$ we have $\lambda'\mathbf{X}_n \Rightarrow \lambda'\mathbf{X}$. Formulate and prove some multi-dimensional Central Limit Theorem for independent but not identically distributed random vectors. Hint: use some formulation of one-dimensional Linderberg-Fuller's theorem.

5. Prove the following statements:

   (a) If $X_n = O_p(n^{-\delta})$ for some $\delta > 0$ then $X_n = o_p(1)$;

   (b) If $X_n = o_p(b_n)$ then $X_n = O_p(b_n)$;

   (c) If $X_n = O_p(n^\alpha)$ and $Y_n = O_p(n^\beta)$, then $X_nY_n = O_p(n^{\alpha+\beta})$ and $X_n + Y_n = O_p(\max\{n^\alpha, n^\beta\})$;

   (d) If $X_n = O_p(n^\alpha)$ and $Y_n = o_p(n^\beta)$, then $X_nY_n = o_p(n^{\alpha+\beta})$.