Statistics 330b/600b, spring 2010
Homework # 8
Due: Thursday 1 April

Throughout this sheet $\{X_n : n \in \mathbb{N}\}$ *will be a fixed exchangeable sequence of random variables, each* X_i *taking values in* $\{0,1\}$*. That is, for each* n*, the probability* $\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\}$ *for* $x_i \in \{0,1\}$ *only depends on* $\sum_{i \le n} x_i$*. Equivalently, for each permutation* π *of* $1, 2, \dots, n$*, the random vectors*

$$(X_1, X_2, \dots, X_n) \qquad \textit{and} \qquad ((X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)})$$

have the same joint distribution.

Define $S_n := \sum_{i \le n} X_i$ *and* $\mathcal{G}_n := \sigma\{S_n, S_{n+1}, \dots\}$ *and* $\mathcal{G}_\infty := \cap_{n \in \mathbb{N}} \mathcal{G}_n$*.*

*[1] Show that $\dots, (S_n/n, \mathcal{G}_n), \dots, (S_1, \mathcal{G}_1)$ is a martingale, by these steps.

(i) For each bounded, $\mathcal{B}(\mathbb{R}^k)$-measurable function f, show that

$$\mathbb{P}X_j f(S_n, S_{n+1}, \dots, S_{n+k}) = \mathbb{P}X_1 f(S_n, S_{n+1}, \dots, S_{n+k}) \qquad \text{for } j = 1, 2, \dots, n.$$

Hint: Think of the integrand on the right-hand side as $g(X_1, X_2, \dots, X_{n+k})$. What happens if you interchange X_1 and X_j?

(ii) With f as in part (i), show that

$$\mathbb{P}S_n f(S_n, S_{n+1}, \dots, S_{n+k}) = n\mathbb{P}X_1 f(S_n, S_{n+1}, \dots, S_{n+k})$$
$$\mathbb{P}S_{n-1} f(S_n, S_{n+1}, \dots, S_{n+k}) = (n-1)\mathbb{P}X_1 f(S_n, S_{n+1}, \dots, S_{n+k})$$

(iii) Use some sort of generating class argument to establish the asserted martingale property.

[2] Explain why there exists a $\mathcal{G}_\infty$-measurable random variable Z, taking values in $[0,1]$, for which $S_n/n \to Z$ almost surely.

*[3] For each subset J of $\{1, 2, \dots, n\}$ write X_J for $\prod_{i \in J} X_i$ and $|J|$ for the size of J.

(i) For a fixed positive integer m, show that there exist constants C_j for $j = 1, 2, \dots, m$, with $C_m = m!$, such that

$$S_n^m = \sum\nolimits_{j=1}^m C_j \sum\nolimits_{|J|=j} X_J \qquad \text{for each } n \ge m.$$

(ii) Deduce that

$$\sum\nolimits_{j=1}^m C_j \binom{n}{j} \mathbb{P}(X_1 \dots X_j \mid S_n = k) = k^m$$

(iii) Deduce that $\mathbb{P}(X_1 \dots X_m \mid S_n = k) = (k/n)^m + O(1/n)$, with the error bound holding uniformly over $k = 0, 1, \dots, n$.

*[4] Write Q for the distribution of the random variable Z from Problem [2].

(i) For each bounded, continuous function f on $[0,1]$ explain why

$$\mathbb{P}X_1 \dots X_m f(Z) = \lim_{n\to\infty} \mathbb{P}X_1 \dots X_m f(S_n/n) = Q\theta^m f(\theta)$$

(ii) You may assume that the conditional probability distribution $\mathbb{P}_\theta(\cdot) = \mathbb{P}(\cdot \mid Z = \theta)$ exists. (It does.) Show that $\mathbb{P}_\theta(X_1 \dots X_m) = \theta^m$ a.e. $[Q]$.

(iii) Deduce that

$$\mathbb{P}_\theta X_1 X_2 \dots X_m (1 - X_{m+1}) \dots (1 - X_{m+k}) = \theta^m (1-\theta)^k$$

(iv) Explain why Q is uniquely determined by $Q\theta^m$ for $m = 1, 2, \dots$. Hint: Weierstrass approximation theorem.

[5] For the Polya urn model with the urn initially containing r red balls and b black balls, and with $d = 1$, show that

$$\mathbb{P}(X_1 \dots X_m) = \prod_{i=0}^{m-1} \frac{r+i}{r+b+i} = Q\theta^m,$$

where Q is the Beta(α, β) distribution for some choice of α and β. What does this tell you about the limiting distribution of S_n/n?