# Exercises 4, Stat 206B/Math 223B, Spring 2011

1. Let $X_1, \ldots, X_n$ i.i.d. with $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$, and for some $k$ with $1 \leqslant k \leqslant n$, define

$$Y_j = (1 - X_{j-1}) \prod_{i=j}^{j+k-1} X_i, \; j = 1, \ldots, n,$$

where the bounds of the product are defined modularly (e.g. $Y_n = (1-X_{n-1})X_n X_1 \cdots X_{k-1}$). Let

$$W = \sum_{j=1}^{n} Y_j$$

and let $d_{\mathrm{TV}}$ denote the total variation distance.

(a) Using the method of dependency graphs, find a bound on $d_{\mathrm{TV}}(W, Z)$, where $Z$ has a Poisson distribution with mean $\mathbb{E}[W]$.

(b) Find the regime of $n$ and $k \equiv k(n)$ where $d_{\mathrm{TV}}(W, Z)$ goes to zero as $n$ goes to infinity and $\mathbb{E}[W]$ is bounded away from zero and infinity.

(c) Repeat parts (a) and (b) using a size-biased coupling and compare your bounds to the method of dependency graphs.

2. Let $\sigma$ be a uniformly chosen random permutation of $\{1, \ldots, n\}$, and for fixed $1 \leqslant k \leqslant n$, let $W$ be the size of the set $\{i : i \leqslant \sigma_i < i + k\}$ (addition is modular). Note that if $k = 1$, then $W$ is the number of fixed points of $\sigma$. Find the mean and variance of $W$ and find an upper bound on the total variation distance between $W$ and a Poisson random variable with the same mean as $W$. If you are having trouble solving this problem, you may refer to Chapter 4 of "Poisson Approximation" by Barbour, Holst, and Janson where a solution is worked out. You may use their work as a guide, but please do not use any of their out-the-door results.

3. Let $W$ be an integer valued random variable with mean $\mu$ and variance $\sigma^2$. If $\tilde{W} := (W - \mu)/\sigma$ has approximately a normal distribution, it may be the case that $W$ is close in the total variation distance to some discrete analog of the normal distribution. This exercise adapts Stein's method of exchangeable pairs for Poisson approximation to such an analog - the *translated Poisson* distribution.

   Let $s = \lfloor \mu - \sigma^2 \rfloor$ and $\gamma = \mu - \sigma^2 - s$, where $\lfloor x \rfloor$ is the greatest integer no larger than $x$. We say the random variable $Z$ has the translated Poisson distribution with parameters $\mu$ and $\sigma^2$ if $Z - s$ is Poisson distributed with mean $\sigma^2 + \gamma$.

(a) Retaining the notation and definitions above, let $(W-\mu, W'-\mu)$ be an $a$-Stein pair and also let $W'-W \in \{-1,0,1\}$. Show that

$$d_{\mathrm{TV}}(W,Z) \leqslant \frac{\sqrt{\mathrm{Var}\left(\mathbb{P}(W'=W+1|W)\right)}}{a\sigma^2} + \frac{2}{\sigma^2}. \tag{1}$$

*Hint:* Total variation distance is translation invariant.

(b) Let $W=\sum_{i=1}^{n} X_i$ where $\mathbb{P}(X_i=1)=1-\mathbb{P}(X_i=0)=p_i$ and the summands are independent. Use (1) to find an upper bound on the total variation distance between $W$ and an appropriate translated Poisson distribution that tends to zero as $\mathrm{Var}(W)$ tends to infinity.

(c) Show that a non-negative integer valued random variable $W$ with positive variance has a representation as a sum of independent indicators if and only if its generating function $G(t)=\sum_{k\geqslant 0}\mathbb{P}(W=k)t^k$ is a polynomial with real roots.

Remarks: These last two items imply that if a non-negative integer-valued random variable has a polynomial generating function with real roots, then (1) can be applied and will yield a good approximation provided that $\mathrm{Var}(W)$ is large. An example where this is useful is the number of descents of a random permutation. Also note that the anti-voter model studied in lecture fits into the framework where (1) can be applied so that we get approximation bounds in the stronger total variation metric for free from the analysis already performed.