Probability approximations using Stein's method

Tutorial 2

1. **Stein's method for geometric approximation, part II:** Let $Z \sim \text{Geom}(p)$ and let W be a non-negative, integer-valued random variable with $\mathbb{P}(W = 0) = p$. Define a random variable V which has the same distribution as $(W - 1|W > 0)$.

   (a) Show that
   $$d_{TV}(W, Z) \leq \frac{1-p}{p}\mathbb{P}(W \neq V)\,.$$

   (b) Let $N \sim \text{Geom}(p)$ and let $X, X_1, X_2, \ldots$ be a sequence of IID positive, integer-valued random variables independent of N. Define $W = X_1 + \cdots + X_N$. Such *geometric sums* arise in a variety of applications in queuing theory, insurance, the analysis of Markov chains and other applications. Show that with this choice of W we may let $V = W + X - 1$, and hence that
   $$d_{TV}(W, Z) \leq \frac{1-p}{p}\mathbb{P}(X > 1)\,.$$

2. Let $X_1, \ldots, X_n$ be indicator random variables with $\mathbb{P}(X_i = 1) = p_i$. Let $\lambda = p_1 + \cdots + p_n$ and $Z \sim \text{Po}(\lambda)$. For each $i = 1, \ldots, n$, let U_i have the same distribution as W and let V_i have the same distribution as $(W - 1|X_i = 1)$. Suppose that we can construct U_i, V_i and X_i on the same probability space such that $U_i - X_i \leq V_i$ with probability 1 for each $i = 1, \ldots, n$. Show that
   $$d_{TV}(W, Z) \leq \frac{1}{\lambda}\left(\text{Var}(W) - \lambda + 2\sum_{i=1}^{n} p_i^2\right)\,.$$

3. Suppose m people each have a birthday on one of n days of the year, where each person's birthday is chosen uniformly at random (independently of everyone else) from the n available days. Let Y_i be the number of these m people born on day i, for $i = 1, \ldots, n$. Let $X_i = I(Y_i = 0)$, and $W = X_1 + \cdots + X_n$ count the number of days on which no-one has a birthday.

   (a) Show that $\mathbb{E}[W] = n(1 - 1/n)^m$.

   (b) For a given $i \in \{1, \ldots, n\}$, consider the setting in which the Y_i people born on day i have their birthday reassigned uniformly at random (and independently of all else) to one of the remaining $n-1$ days. With this reassignment, let $1+V_i$ denote the number of days on which no-one has a birthday. Show that

$$\mathbb{E}[V_i] = (n-1)\left(1 - \frac{1}{n-1}\right)^m .$$

4. Let $Z \sim \mathrm{N}(0,1)$ and let $h : \mathbb{R} \to \mathbb{R}$ be a measurable function with $\mathbb{E}|h(Z)| < \infty$. By using an appropriate integrating factor, or otherwise, show that a solution to the equation

$$h(x) - \mathbb{E}h(Z) = f'(x) - xf(x) \tag{1}$$

is given by

$$\begin{aligned} f(x) &= e^{x^2/2} \int_{-\infty}^{x} (h(y) - \mathbb{E}h(Z))\, e^{-y^2/2}\, dy \\ &= -e^{x^2/2} \int_{x}^{\infty} (h(y) - \mathbb{E}h(Z))\, e^{-y^2/2}\, dy \,, \end{aligned}$$

and state the form of the general solution to the equation (1).

5. Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}X_i = 0$ and $\sum_{i=1}^n \mathrm{Var}(X_i) = 1$. Let $W = X_1 + \cdots + X_n$. Show that

$$\left| \mathbb{E}|W| - \sqrt{\frac{2}{\pi}} \right| \le 4 \sum_{i=1}^{n} \mathbb{E}|X_i^3| \,.$$