2.7 Exercises

1. If $X \sim$ Poisson(a) and $Y \sim$ Poisson(b), with $b > a$, use coupling to show that $Y \geq_{st} X$.

2. Suppose a particle starts at position 5 on a number line and at each time period the particle moves one position to the right with probability p and, if the particle is above position 0, moves one position to left with probability $1 - p$. Let $X_n(p)$ be the position of the particle at time n for the given value of p. Use coupling to show that $X_n(a) \geq_{st} X_n(b)$ for any n if $a \geq b$.

3. Let X, Y be indicator variables with $E[X] = a$ and $E[Y] = b$. (a) Show how to construct a maximal coupling $\widehat{X}, \widehat{Y}$ for X and Y, and then compute $P(\widehat{X} = \widehat{Y})$ as a function of a, b. (b) Show how to construct a minimal coupling to minimize $P(\widehat{X} = \widehat{Y})$.

4. In a room full of n people, let X be the number of people who share a birthday with at least one other person in the room. Then let Y be the number of pairs of people in the room having the same birthday. (a) Compute $E[X]$ and $Var(X)$ and $E[Y]$ and $Var(Y)$. (b) Which of the two variables X or Y do you believe will more closely follow a Poisson distribution? Why? (c) In a room of 51 people, it turns out there are 3 pairs with the same birthday and also a triplet (3 people) with the same birthday. This is a total of 9 people and also 6 pairs. Use a Poisson approximation to estimate $P(X > 9)$ and $P(Y > 6)$. Which of these two approximations do you think will be better? Have we observed a rare event here?

5. Compute a bound on the accuracy of the better approximation in the previous exercise part (c) using the Stein-Chen method.

6. For discrete X, Y prove $d_{TV}(X, Y) = \frac{1}{2}\sum_x |P(X = x) - P(Y = x)|$

7. For discrete X, Y show that $P(X \neq Y) \geq d_{TV}(X, Y)$ and show also that there exists a coupling that yields equality.

8. Compute a bound on the accuracy of a normal approximation for a Poisson random variable with mean 100.

owing to the inequality $\lambda^k/k! \geq \lambda^k/[(k-1)!s]$ for $k = 1, \ldots, s$. For $s > j$ again the difference $g_j(s+1) - g_j(s) \leq 0$ because

$$\frac{s}{\lambda} \sum_{k=s+1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \quad \leq \quad \sum_{k=s}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!}$$

owing to the opposite inequality $\lambda^k/k! \leq \lambda^k/[(k-1)!s]$ for $k \geq s+1$. Only the difference $g_j(j+1) - g_j(j) \geq 0$, and this difference is bounded above by

$$\begin{aligned}
g_j(j+1) - g_j(j) &= \frac{1}{\lambda} \sum_{k=j+1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} + \frac{1}{j} \sum_{k=0}^{j-1} e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \frac{e^{-\lambda}}{\lambda} \left[\sum_{k=j+1}^{\infty} \frac{\lambda^k}{k!} + \sum_{k=1}^{j} \frac{\lambda^k}{k!} \frac{k}{j} \right] \\
&\leq \frac{e^{-\lambda}}{\lambda} (e^{\lambda} - 1) \\
&= \frac{1 - e^{-\lambda}}{\lambda}.
\end{aligned}$$

This upper-bound inequality carries over to

$$g(s+1) - g(s) \quad = \quad \sum_{j \in A} [g_j(s+1) - g_j(s)]$$

since only one difference on its right-hand sum is nonnegative for any given s. Finally, inspection of the solution (12.7) makes it evident that the function $h(s) = -g(s)$ solves the difference equation (12.6) with the complement A^c replacing A. It follows that

$$g(s) - g(s+1) \quad = \quad h(s+1) - h(s) \quad \leq \quad \frac{1 - e^{-\lambda}}{\lambda},$$

and this completes the proof that $g(s)$ satisfies the Lipschitz condition.

12.5 Problems

1. For a random permutation $\sigma_1, \ldots, \sigma_n$ of $\{1, \ldots, n\}$, let $X_\alpha = 1_{\{\sigma_\alpha = \alpha\}}$ be the indicator of a match at position α. Show that the total number of matches $S = \sum_{\alpha=1}^{n} X_\alpha$ satisfies the coupling bound

$$\|\pi_S - \pi_Z\|_{TV} \quad \leq \quad \frac{2(1 - e^{-1})}{n},$$

where Z follows a Poisson distribution with mean 1.

2. In the ménage problem, prove that $\text{Var}(S) = 2 - 2/(n-1)$.

3. In certain situations the hypergeometric distribution can be approximated by a Poisson distribution. Suppose that w white balls and b black balls occupy a box. If you extract $n < w + b$ balls at random, then the number of white balls S extracted follows a hypergeometric distribution. Note that if we label the white balls $1, \ldots, w$, and let X_α be the random variable indicating whether white ball α is chosen, then $S = \sum_{\alpha=1}^{w} X_\alpha$. One can construct a coupling between S and V_α by the following device. If white ball α does not show up, then randomly take one of the balls extracted and exchange it for white ball α. Calculate an explicit Chen-Stein bound, and give conditions under which the Poisson approximation to S will be good.

4. In the context of Example 12.3.1 on the law of rare events, prove the less stringent bound

$$\|\pi_S - \pi_Z\|_{TV} \quad \leq \quad \sum_{\alpha=1}^{n} p_\alpha^2 \qquad .$$

by invoking Problems 14 and 15 of Chapter 7.

5. Consider the n-dimensional unit cube $[0,1]^n$. Suppose that each of its $n2^{n-1}$ edges is independently assigned one of two equally likely orientations. Let S be the number of vertices at which all neighboring edges point toward the vertex. The Chen-Stein method implies that S has an approximate Poisson distribution Z with mean 1. Use the neighborhood method to verify the estimate

$$\|\pi_S - \pi_Z\|_{TV} \leq (n+1)2^{-n}(1 - e^{-1}).$$

(Hints: Let I be the set of all 2^n vertices, X_α the indicator that vertex α has all of its edges directed toward α, and $N_\alpha = \{\beta : \|\beta - \alpha\| \leq 1\}$. Note that X_α is independent of those X_β with $\|\beta - \alpha\| > 1$. Also, $p_{\alpha\beta} = 0$ for $\|\beta - \alpha\| = 1$.)

6. A graph with n nodes is created by randomly connecting some pairs of nodes by edges. If the connection probability per pair is p, then all pairs from a triple of nodes are connected with probability p^3. For p small and $\lambda = \binom{n}{3}p^3$ moderate in size, the number of such triangles in the random graph is approximately Poisson with mean λ. Use the neighborhood method to estimate the total variation error in this approximation.

7. Suppose n balls (people) are uniformly and independently distributed into m boxes (days of the year). The birthday problem involves finding the approximate distribution of the number of boxes that receive