Stat 110 Strategic Practice 11, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 Law of Large Numbers, Central Limit Theorem

1. Give an intuitive argument that the Central Limit Theorem implies the Weak Law of Large Numbers, without worrying about the different forms of convergence; then briefly explain how the forms of convergence involved are different.

2. (a) Explain why the Pois(n) distribution is approximately Normal if n is a large positive integer (specifying what the parameters of the Normal are).

   (b) Stirling's formula is an amazingly accurate approximation for factorials:

   $$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

   where in fact the ratio of the two sides goes to 1 as $n \to \infty$. Use (a) to give a quick heuristic derivation of Stirling's formula by using a Normal approximation to the probability that a Pois(n) r.v. is n, with the continuity correction: first write $P(N = n) = P(n - \frac{1}{2} < N < n + \frac{1}{2})$, where $N \sim \text{Pois}(n)$.

3. Let $T_1, T_2, \dots$ be i.i.d. Student-t r.v.s with $m \geq 3$ degrees of freedom. Find constants a_n and b_n (in terms of m and n) such that $a_n(T_1 + T_2 + \dots + T_n - b_n)$ converges to $\mathcal{N}(0, 1)$ in distribution as $n \to \infty$.

4. Let $X_1, X_2, \dots$ be i.i.d. positive random variables with mean 2. Let $Y_1, Y_2, \dots$ be i.i.d. positive random variables with mean 3. Show that

   $$\frac{X_1 + X_2 + \dots + X_n}{Y_1 + Y_2 + \dots + Y_n} \to \frac{2}{3}$$

   with probability 1. Does it matter whether the X_i are independent of the Y_j?

5. Let f be a complicated function whose integral $\int_a^b f(x)dx$ we want to approximate. Assume that $0 \leq f(x) \leq c$. Let A be the rectangle in the (x, y)-plane given by $a \leq x \leq b$ and $0 \leq y \leq c$. Pick i.i.d. uniform points $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ in the rectangle A.

   How would you use these random points to approximate the integral? (This is an example of a *Monte Carlo method*.) Show that the estimate converges to the true value of the integral as $n \to \infty$. Hint: look at whether each point is in the area below the curve $y = f(x)$.

2 Multivariate Normal

1. Show that if (X_1, X_2, X_3) is Multivariate Normal, then so is the subvector (X_1, X_2).

2. Is it true that if $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_m)$ are Multivariate Normals with $\mathbf{X}$ independent of $\mathbf{Y}$, then the "concatenated" random vector $\mathbf{W} = (X_1, \dots, X_n, Y_1, \dots, Y_m)$ is Multivariate Normal?

3. Let (X, Y) be Bivariate Normal, with $X, Y \sim \mathcal{N}(0, 1)$ (marginally) and correlation ρ, where $-1 < \rho < 1$. Find a, b, c, d (in terms of ρ) such that $Z = aX + bY$ and $W = cX + dY$ are independent $\mathcal{N}(0, 1)$.

4. Let $X_1, \dots, X_n$ be i.i.d. $\mathcal{N}(\mu, \sigma^2)$, with $n \geq 2$. Let $\bar{X}_n$ be the sample mean, and let $S_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$ (this is called the *sample variance*, and is an unbiased estimator for σ^2). Show that $\bar{X}_n$ and S_n^2 are independent by applying MVN ideas and results to the vector $(\bar{X}_n, X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$.

3 Markov Chains

1. Let $X_0, X_1, X_2 \dots$ be a Markov chain. Show that $X_0, X_2, X_4, X_6, \dots$ is also a Markov chain, and explain why this makes sense intuitively.

   Let $Y_n = X_{2n}$; we need to show $Y_0, Y_1, \dots$ is a Markov chain. By the definition of a Markov chain, we know that $X_{2n+1}, X_{2n+2}, \dots$ ("the future" if we define the "present" to be time $2n$) is conditionally independent of $X_0, X_1, \dots, X_{2n-2}, X_{2n-1}$ ("the past"), given X_{2n}. So given Y_n, we have that $Y_{n+1}, Y_{n+2}, \dots$ is conditionally independent of $Y_0, Y_1, \dots, Y_{n-1}$. Thus,

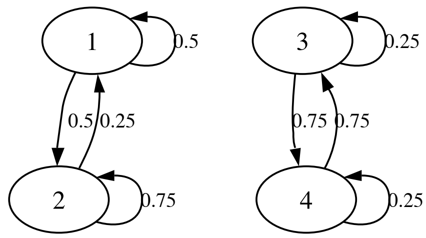
   $$P(Y_{n+1} = y | Y_0 = y_0, \dots, Y_n = y_n) = P(Y_{n+1} = y | Y_n = y_n).$$

2. Consider the Markov chain shown below, with state space $\{1, 2, 3, 4\}$ and where the labels next to arrows indicate the probabilities of those transitions.

   (a) Write down the transition matrix Q for this chain.

   (b) Which states (if any) are recurrent? Which states (if any) are transient?

   (c) Find two different stationary distributions for the chain.

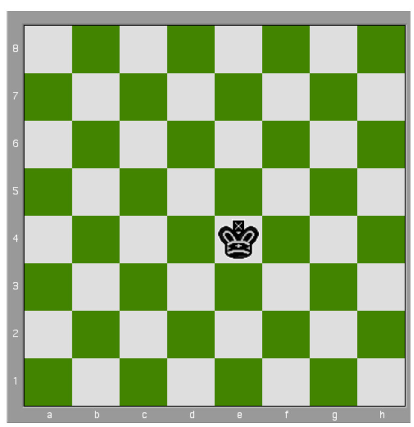


3. A cat and a mouse move independently back and forth between two rooms. At each time step, the cat moves from the current room to the other room with probability 0.8. Starting from Room 1, the mouse moves to Room 2 with probability 0.3 (and remains otherwise). Starting from Room 2, the mouse moves to Room 1 with probability 0.6 (and remains otherwise).

   (a) Find the stationary distributions of the cat chain and of the mouse chain.

   (b) Note that there are 4 possible (cat, mouse) states: both in Room 1, cat in Room 1 and mouse in Room 2, cat in Room 2 and mouse in Room 1, and both in Room 2. Number these cases 1,2,3,4 respectively, and let Z_n be the number of the current (cat, mouse) state at time n. Is $Z_0, Z_1, Z_2, \ldots$ a Markov chain?

4. In chess, the king can move one square at a time in any direction (horizontally, vertically, or diagonally).



   For example, in the diagram, from the current position the king can move to any of 8 possible squares. A king is wandering around on an otherwise empty 8 by 8 chessboard, where for each move all possibilities are equally likely. Find the stationary distribution of this chain (of course, don't list out a vector of length 64 explicitly! Classify the 64 squares into "types" and say what the stationary probability is for a square of each type).

5. Markov chains have recently been applied to code-breaking; this problem will consider one way in which this can be done. A *substitution cipher* is a permutation g of the letters from a to z, where a message is enciphered by replacing each letter α by $g(\alpha)$. For example, if g is the permutation given by

```
abcdefghijklmnopqrstuvwxyz
zyxwvutsrqponmlkjihgfedcba
```

where the second row lists the values $g(a), g(b), \ldots, g(z)$, then we would encipher the word "statistics" as "hgzgrhgrxh". The state space is all $26! \approx 4 \cdot 10^{26}$ permutations of the letters a through z.

(a) Consider the chain that picks two different random coordinates between 1 and 26 and swaps those entries of the 2nd row, e.g., if we pick 7 and 20, then

```
abcdefghijklmnopqrstuvwxyz
zyxwvutsrqponmlkjihgfedcba
```

becomes

```
abcdefghijklmnopqrstuvwxyz
zyxwvugsrqponmlkjihtfedcba
```

Find the probability of going from a permutation g to a permutation h in one step (for all g, h), and find the stationary distribution of this chain.

(b) Suppose we have a system that assigns a positive "score" $s(g)$ to each permutation g (intuitively, this could be a measure of how likely it would be to get the observed enciphered text, given that g was the cipher used). Consider the following Markov chain. Starting from any state g, generate a "proposal" h using the chain from (a). If $s(g) \leq s(h)$, then go to h (i.e., accept the proposal). Otherwise, flip a coin with probability $s(h)/s(g)$ of Heads. If Heads, go to h (i.e., accept the proposal); if Tails, stay at g. Show that this chain is reversible and has stationary distribution proportional to the list of all scores $s(g)$.

Hint: for $g \neq h$, let $q(g, h)$ be the probability of going from g to h in one step, and show that $s(g)q(g, h) = s(h)q(h, g)$. To compute $q(g, h)$, note that it is the probability of proposing h when at g, times the probability of accepting the proposal.

Stat 110 Strategic Practice 11 Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 Law of Large Numbers, Central Limit Theorem

1. Give an intuitive argument that the Central Limit Theorem implies the Weak Law of Large Numbers, without worrying about the different forms of convergence; then briefly explain how the forms of convergence involved are different.

   Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. The CLT says that

   $$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \to \mathcal{N}(0,1)$$

   in distribution. Since $\sqrt{n}$ goes to infinity, it makes sense that $\bar{X}_n - \mu$ should go to 0, to prevent $\sqrt{n}(\bar{X}_n - \mu)$ from exploding. The CLT is more informative in the sense that it gives the shape of the distribution of the sample mean (after standardization), and gives information about the rate at which the sample mean goes to the true mean (replacing $\sqrt{n}$ by a different power of n, the expression would go to 0 or ∞ rather than to a Normal distribution).

   On the other hand, the CLT is a statement about convergence in *distribution* (i.e., the distribution of the r.v. on the left goes to the standard Normal distribution), while the Weak Law of Large Numbers says that the r.v. $\bar{X}_n$ will be extremely close to μ with extremely high probability, for n large enough.

2. (a) Explain why the Pois(n) distribution is approximately Normal if n is a large positive integer (specifying what the parameters of the Normal are).

   Let $S_n = X_1 + \cdots + X_n$, with $X_1, X_2, \ldots$ i.i.d. $\sim$ Pois(1). Then $S_n \sim$ Pois(n) and for n large, S_n is approximately $\mathcal{N}(n, n)$ by the CLT.

   (b) Stirling's formula is an amazingly accurate approximation for factorials:

   $$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

   where in fact the ratio of the two sides goes to 1 as $n \to \infty$. Use (a) to give a quick heuristic derivation of Stirling's formula by using a Normal approximation to the probability that a Pois(n) r.v. is n, with the continuity correction: first write $P(N = n) = P(n - \frac{1}{2} < N < n + \frac{1}{2})$, where $N \sim$ Pois(n).

Let $N \sim \text{Pois}(n)$ and $X \sim \mathcal{N}(n, n)$. Then

$$P(N = n) \approx P\left(n - \frac{1}{2} < X < n + \frac{1}{2}\right) = \frac{1}{\sqrt{2\pi n}} \int_{n-1/2}^{n+1/2} e^{\frac{-(x-n)^2}{2n}} dx.$$

The integral is approximately 1 since the interval of integration has length 1 and for large n the integrand is very close to 1 throughout the interval. So

$$e^{-n} n^n / n! \approx (2\pi n)^{-1/2}.$$

Rearranging this gives exactly Stirling's formula.

3. Let $T_1, T_2, \ldots$ be i.i.d. Student-t r.v.s with $m \geq 3$ degrees of freedom. Find constants a_n and b_n (in terms of m and n) such that $a_n(T_1 + T_2 + \cdots + T_n - b_n)$ converges to $\mathcal{N}(0, 1)$ in distribution as $n \to \infty$.

   First let us find the mean and variance of each T_j. Let $T = \frac{Z}{\sqrt{V/m}}$ with $Z \sim \mathcal{N}(0, 1)$ independent of $V \sim \chi^2_m$. By LOTUS, for $G \sim \text{Gamma}(a, \lambda)$, $E(G^r)$ is $\lambda^{-r}\Gamma(a + r)/\Gamma(a)$ for $r > -a$, and does not exist for $r \leq -a$. So

$$\begin{aligned}
E(T) &= E(Z)E(\frac{1}{\sqrt{V/m}}) = 0, \\
\text{Var}(T) = E(T^2) - (ET)^2 &= mE(Z^2)E\left(\frac{1}{V}\right) \\
&= m\frac{(1/2)\Gamma(m/2 - 1)}{\Gamma(m/2)} \\
&= \frac{m\Gamma(m/2 - 1)}{2\Gamma(m/2)} \\
&= \frac{m/2}{m/2 - 1} = \frac{m}{m - 2}.
\end{aligned}$$

   By the CLT (and linearity of E, and the fact that the variance of the sum of uncorrelated r.v.s is the sum of the variances), this is true for

$$\begin{aligned}
b_n &= E(T_1) + \ldots + E(T_n) = 0, \\
a_n &= \frac{1}{\sqrt{\text{Var}(T_1) + \cdots + \text{Var}(T_n)}} = \sqrt{\frac{m - 2}{mn}}.
\end{aligned}$$

4. Let $X_1, X_2, \ldots$ be i.i.d. positive random variables with mean 2. Let $Y_1, Y_2, \ldots$ be i.i.d. positive random variables with mean 3. Show that
$$\frac{X_1 + X_2 + \cdots + X_n}{Y_1 + Y_2 + \cdots + Y_n} \to \frac{2}{3}$$
with probability 1. Does it matter whether the X_i are independent of the Y_j?

   By the Law of Large Numbers, $\frac{X_1+X_2+\cdots+X_n}{n} \to 2$ with probability 1 and $\frac{Y_1+Y_2+\cdots+Y_n}{n} \to 3$ with probability 1, as $n \to \infty$. Note that if two events A and B both have probability 1, then the event $A \cap B$ also has probability 1. So with probability 1, *both* the convergence involving the X_i and the convergence involving the Y_j occur. Therefore,
$$\frac{X_1 + X_2 + \cdots + X_n}{Y_1 + Y_2 + \cdots + Y_n} = \frac{(X_1 + X_2 + \cdots + X_n)/n}{(Y_1 + Y_2 + \cdots + Y_n)/n} \to \frac{2}{3} \text{ with probability 1}$$
as $n \to \infty$. It was not necessary to assume that the X_i are independent of the Y_j because of the pointwise with probability 1 convergence.

5. Let f be a complicated function whose integral $\int_a^b f(x)dx$ we want to approximate. Assume that $0 \leq f(x) \leq c$. Let A be the rectangle in the (x, y)-plane given by $a \leq x \leq b$ and $0 \leq y \leq c$. Pick i.i.d. uniform points $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ in the rectangle A.

   How would you use these random points to approximate the integral? (This is an example of a *Monte Carlo method.*) Show that the estimate converges to the true value of the integral as $n \to \infty$. Hint: look at whether each point is in the area below the curve $y = f(x)$.

   Let B be the region under the curve $y = f(x)$ (and above the x axis) for $a \leq x \leq b$, so the desired integral is the area of region B. Define indicator r.v.s $I_1, \ldots, I_n$ by letting $I_j = 1$ if X_j is in B and $I_j = 0$ otherwise. Let $\mu = E(I_1)$. In terms of areas, μ is the ratio of the area of B to the area of the rectangle A:
$$\mu = E(I_j) = P(I_j = 1) = \frac{\int_a^b f(x)dx}{c(b-a)}.$$
   We can estimate μ using $\frac{1}{n}\sum_{j=1}^n I_j$, and then estimate the desired integral by
$$\int_a^b f(x)dx \approx c(b-a)\frac{1}{n}\sum_{j=1}^n I_j.$$
   Since the I_j are i.i.d. with mean μ, it follows from the Law of Large Numbers that with probability 1, the estimate converges to the true value of the integral.

2 Multivariate Normal

1. Show that if (X_1, X_2, X_3) is Multivariate Normal, then so is the subvector (X_1, X_2).

   Any linear combination $t_1X_1 + t_2X_2$ can be thought of as a linear combination of X_1, X_2, X_3 (where the coefficient of X_3 is 0), so $t_1X_1 + t_2X_2$ is Normal for all t_1, t_2, which shows that (X_1, X_2) is MVN.

2. Is it true that if $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ are Multivariate Normals with $\mathbf{X}$ independent of $\mathbf{Y}$, then the "concatenated" random vector $\mathbf{W} = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ is Multivariate Normal?

   Yes, since any linear combination $s_1X_1 + \cdots + s_nX_n + t_1Y_1 + \cdots + t_mY_m$ is Normal, because $s_1X_1 + \cdots + s_nX_n$ and $t_1Y_1 + \cdots + t_mY_m$ are Normal (by definition of $\mathbf{X}$ and $\mathbf{Y}$ being MVN) and are independent, so their sum is Normal.

3. Let (X, Y) be Bivariate Normal, with $X, Y \sim \mathcal{N}(0, 1)$ (marginally) and correlation ρ, where $-1 < \rho < 1$. Find a, b, c, d (in terms of ρ) such that $Z = aX + bY$ and $W = cX + dY$ are independent $\mathcal{N}(0, 1)$.

   We need to find a, b, c, d such that

$$\begin{aligned}
\mathrm{Cov}(Z, W) &= \mathrm{Cov}(aX + bY, cX + dY) \\
&= \mathrm{Cov}(aX, cX) + \mathrm{Cov}(bY, cX) + \mathrm{Cov}(aX, dY) + \mathrm{Cov}(bY, dY) \\
&= ac\mathrm{Var}(X) + (bc + ad)\mathrm{Cov}(X, Y) + db\mathrm{Var}(Y) \\
&= ac + (bc + ad)\rho + db = 0, && (1) \\
\mathrm{Var}(Z) &= \mathrm{Cov}(aX + bY, aX + bY) \\
&= a^2\mathrm{Var}(X) + 2ab\mathrm{Cov}(X, Y) + b^2\mathrm{Var}(Y) \\
&= a^2 + 2ab\rho + b^2 = 1, && (2) \\
\mathrm{Var}(W) &= c^2 + 2cd\rho + d^2 = 1. && (3)
\end{aligned}$$

   Since uncorrelated implies independent within a Multivariate Normal, if we can solve these equations then Z and W are as desired. These are 3 equations in 4 unknowns, so there may be many solutions; let us simplify by looking for a solution with $a = 1$. Then by (2), $2b\rho + b^2 = b(2\rho + b) = 0$, so $b = 0$ or $b = -2\rho$. Trying the possibility $b = 0$, we then have $c = -d\rho$ by (1). Plugging this into equation (3) gives us

$$(-d\rho)^2 + 2(-d\rho)d\rho + d^2 = d^2(1 - \rho^2) = 1,$$

which yields $d = \pm 1/\sqrt{1-\rho^2}$. So if we let $a = 1, b = 0, c = -\rho/\sqrt{1-\rho^2}$, and $d = 1/\sqrt{1-\rho^2}$, then $Z = aX + bY$ and $W = cX + dY$ are i.i.d. $\mathcal{N}(0,1)$.

4. Let $X_1, \dots, X_n$ be i.i.d. $\mathcal{N}(\mu, \sigma^2)$, with $n \geq 2$. Let $\bar{X}_n$ be the sample mean, and let $S_n^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \bar{X}_n)^2$ (this is called the *sample variance*, and is an unbiased estimator for σ^2). Show that $\bar{X}_n$ and S_n^2 are independent by applying MVN ideas and results to the vector $(\bar{X}_n, X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$.

   The vector $\bar{X}_n, X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ is MVN since any linear combination of its components can be written as a linear combination of $X_1, \dots, X_n$. Now compute the covariance of $\bar{X}_n$ with $X_j - \bar{X}_n$:

   $$\text{Cov}(\bar{X}_n, X_j - \bar{X}_n) = E(\bar{X}_n(X_j - \bar{X}_n)).$$

   This can be found by expanding it out, but a faster way is to use Adam's Law to write it as

   $$E(E(\bar{X}_n(X_j - \bar{X}_n)|\bar{X}_n)) = E(\bar{X}_n(E(X_j - \bar{X}_n)|\bar{X}_n)).$$

   But this is 0 since $E(X_j - \bar{X}_n|\bar{X}_n) = E(X_j|\bar{X}_n) - E(\bar{X}_n|\bar{X}_n) = \bar{X}_n - \bar{X}_n = 0$, using the idea of Problem 2 from the Penultimate Homework. Therefore, $\bar{X}_n$ is uncorrelated with $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ (i.e., it is uncorrelated with each component of this vector). Since within a MVN it is true that uncorrelated implies independent, we have that $\bar{X}_n$ is independent of $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$. Thus, $\bar{X}_n$ is independent of S_n^2.

3 Markov Chains

1. Let $X_0, X_1, X_2 \dots$ be a Markov chain. Show that $X_0, X_2, X_4, X_6, \dots$ is also a Markov chain, and explain why this makes sense intuitively.

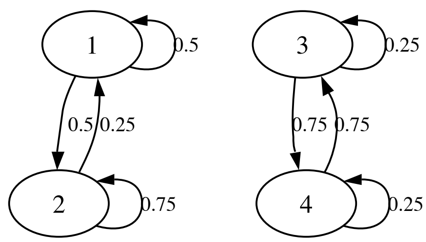
   Let $Y_n = X_{2n}$; we need to show $Y_0, Y_1, \dots$ is a Markov chain. By the definition of a Markov chain, we know that $X_{2n+1}, X_{2n+2}, \dots$ ("the future" if we define the "present" to be time $2n$) is conditionally independent of $X_0, X_1, \dots, X_{2n-2}, X_{2n-1}$ ("the past"), given X_{2n}. So given Y_n, we have that $Y_{n+1}, Y_{n+2}, \dots$ is conditionally independent of $Y_0, Y_1, \dots, Y_{n-1}$. Thus,

   $$P(Y_{n+1} = y|Y_0 = y_0, \dots, Y_n = y_n) = P(Y_{n+1} = y|Y_n = y_n).$$

To find the transition probabilities, call $X_0, X_1, \dots$ the "X-chain" and $Y_0, Y_1, \dots$ the "Y-chain". To find the transition probabilities of the Y-chain, let Q be the transition matrix of the X-chain (labeling the states as $1, 2, \dots, M$). Then Q^2 gives the "two-step" transition probabilities of the X-chain, so $P(Y_{n+1} = j|Y_n = i)$ is the (i, j) entry of Q^2.

This makes sense intuitively since, for example, in predicting what X_{110} will be based on $X_0, X_2, \dots, X_{108}$, we only need to use the most recent value, X_{108}, and then the one-step transitions of the Y-chain correspond to two-step transitions of the X-chain.

2. Consider the Markov chain shown below, with state space $\{1, 2, 3, 4\}$ and where the labels next to arrows indicate the probabilities of those transitions.



(a) Write down the transition matrix Q for this chain.

The transition matrix is

$$Q = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 0 & 0 & 0.25 & 0.75 \\ 0 & 0 & 0.75 & 0.25 \end{pmatrix}$$

(b) Which states (if any) are recurrent? Which states (if any) are transient?

All of the states are recurrent. Starting at state 1, the chain will go back and forth between states 1 and 2 forever (sometimes lingering for a while). Similarly, for any starting state, the probability is 1 of returning to that state.

(c) Find two different stationary distributions for the chain.

Solving

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{pmatrix} = \begin{pmatrix} a & b \end{pmatrix}$$

$$(c \quad d) \begin{pmatrix} 0.25 & 0.75 \\ 0.75 & 0.25 \end{pmatrix} = (c \quad d)$$

shows that $(a, b) = (1/3, 2/3)$, and $(c, d) = (1/2, 1/2)$ are stationary distributions on the 1, 2 chain and on the 3, 4 chain respectively, viewed as separate chains. It follows that $(1/3, 2/3, 0, 0)$ and $(0, 0, 1/2, 1/2)$ are both stationary for Q (as is any mixture $p(1/3, 2/3, 0, 0) + (1 - p)(0, 0, 1/2, 1/2)$ with $0 \leq p \leq 1$).

3. A cat and a mouse move independently back and forth between two rooms. At each time step, the cat moves from the current room to the other room with probability 0.8. Starting from Room 1, the mouse moves to Room 2 with probability 0.3 (and remains otherwise). Starting from Room 2, the mouse moves to Room 1 with probability 0.6 (and remains otherwise).

   (a) Find the stationary distributions of the cat chain and of the mouse chain.

   For the cat Markov chain, the stationary distribution is $(1/2, 1/2)$ by symmetry. For the mouse Markov chain: solving $\mathbf{s}Q = \mathbf{s}$ and normalizing yields $(2/3, 1/3)$.

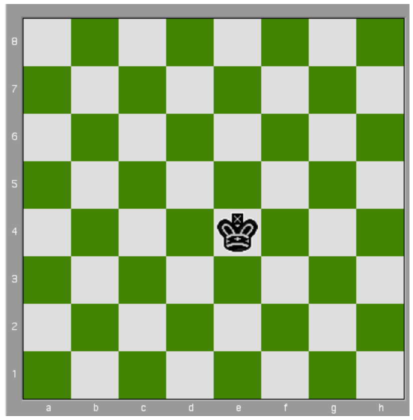
   (b) Note that there are 4 possible (cat, mouse) states: both in Room 1, cat in Room 1 and mouse in Room 2, cat in Room 2 and mouse in Room 1, and both in Room 2. Number these cases 1,2,3,4 respectively, and let Z_n be the number of the current (cat, mouse) state at time n. Is $Z_0, Z_1, Z_2, \ldots$ a Markov chain?

   Yes, it is a Markov chain. Given the current (cat, mouse) state, the past history of where the cat and mouse were previously are irrelevant for computing the probabilities of what the next state will be.

4. In chess, the king can move one square at a time in any direction (horizontally, vertically, or diagonally).

   For example, in the diagram, from the current position the king can move to any of 8 possible squares. A king is wandering around on an otherwise empty 8 by 8 chessboard, where for each move all possibilities are equally likely. Find the stationary distribution of this chain (of course, don't list out a vector of length 64 explicitly! Classify the 64 squares into "types" and say what the stationary probability is for a square of each type).

   There are 4 corner squares, 24 edge squares, and 36 normal squares, where by "edge" we mean a square in the first or last row or column, excluding the 4 corners, and by "normal" we mean a square that's not on the edge or in a



corner. View the chessboard as an undirected network, where there is an edge between two squares if the king can walk from one to the other in one step.

The stationary probabilities are proportional to the degrees. Each corner square has degree 3, each edge square has degree 5, and each normal square has degree 8. The total degree is $420 = 3 \cdot 4 + 24 \cdot 5 + 36 \cdot 8$ (which is also twice the number of edges in the network). Thus, the the stationary probability is $\frac{3}{420}$ for a corner square, $\frac{5}{420}$ for an edge square, and $\frac{8}{420}$ for a normal square.

5. Markov chains have recently been applied to code-breaking; this problem will consider one way in which this can be done. A *substitution cipher* is a permutation g of the letters from a to z, where a message is enciphered by replacing each letter α by $g(\alpha)$. For example, if g is the permutation given by

```
abcdefghijklmnopqrstuvwxyz
zyxwvutsrqponmlkjihgfedcba
```

where the second row lists the values $g(a), g(b), \ldots, g(z)$, then we would encipher the word "statistics" as "hgzgrhgrxh". The state space is all $26! \approx 4 \cdot 10^{26}$ permutations of the letters a through z.

(a) Consider the chain that picks two different random coordinates between 1 and 26 and swaps those entries of the 2nd row, e.g., if we pick 7 and 20, then

```
abcdefghijklmnopqrstuvwxyz
zyxwvutsrqponmlkjihgfedcba
```

becomes

```
abcdefghijklmnopqrstuvwxyz
zyxwvugsrqponmlkjihtfedcba
```

Find the probability of going from a permutation g to a permutation h in one step (for all g, h), and find the stationary distribution of this chain.

The probability of going from g to h (in one step) is 0 unless h can be obtained from g by swapping 2 entries of the second row. Assuming h can be obtained in this way, the probability is $\frac{1}{\binom{26}{2}}$, since there are $\binom{26}{2}$ such swaps, all equally likely.

This Markov chain is irreducible, since by performing enough swaps we can get from any permutation to any other permutation (imagine rearranging a deck of cards by swapping cards two at a time: it is possible to reorder the cards in any desired configuration by doing this enough times). Note that $p(g, h) = p(h, g)$, where $p(g, h)$ is the transition probability of going from g to h. So the chain is reversible with respect to the uniform distribution over all permutations. Thus, the stationary distribution is the uniform distribution over all 26! permutations of the letters a through z.

(b) Suppose we have a system that assigns a positive "score" $s(g)$ to each permutation g (intuitively, this could be a measure of how likely it would be to get the observed enciphered text, given that g was the cipher used). Consider the following Markov chain. Starting from any state g, generate a "proposal" h using the chain from (a). If $s(g) \leq s(h)$, then go to h (i.e., accept the proposal). Otherwise, flip a coin with probability $s(h)/s(g)$ of Heads. If Heads, go to h (i.e., accept the proposal); if Tails, stay at g. Show that this chain is reversible and has stationary distribution proportional to the list of all scores $s(g)$.

Hint: for $g \neq h$, let $q(g, h)$ be the probability of going from g to h in one step, and show that $s(g)q(g, h) = s(h)q(h, g)$. To compute $q(g, h)$, note that it is the probability of proposing h when at g, times the probability of accepting the proposal.

We will check the reversibility condition $s(g)q(g, h) = s(h)q(h, g)$. Assume $g \neq h$ and $q(g, h) \neq 0$ (if $g = h$ or $q(g, h) = 0$, then it follows immediately that the reversibility condition holds). Let $p(g, h)$ be the transition probability from (a) (which is the probability of "proposing" h when at g). First consider the case that $s(g) \leq s(h)$. Then $q(g,h) = p(g,h)$ and

$$q(h,g) = p(h,g)\frac{s(g)}{s(h)} = p(g,h)\frac{s(g)}{s(h)} = q(g,h)\frac{s(g)}{s(h)},$$

so $s(g)q(g,h) = s(h)q(h,g)$. Now consider the case that $s(h) < s(g)$. By a symmetrical argument (reversing the roles of g and h), we again have $s(g)q(g,h) = s(h)q(h,g)$. Thus, the stationary probability of g is proportional to the score $s(g)$.

The algorithm here is an example of the *Metropolis algorithm*, which is a powerful, general method of transmogrifying one Markov chain into a new Markov chain with a certain desired stationary distribution. This algorithm was named one of the ten most important algorithms created in the 20th century!

Stat 110 Ultimate Homework, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. Each day, a very volatile stock rises 70% or drops 50% in price, with equal probabilities (with different days independent). Let X_n be the stock price after n days, starting from an initial value of $X_0 = 100$.

(a) Explain why $\log X_n$ is approximately Normal for n large, and state its parameters.

(b) What happens to $E(X_n)$ as $n \to \infty$?

(c) Use the Law of Large Numbers to find out what happens to X_n as $n \to \infty$.

Hint: let U_n be the number of days the stock rises up until time n.

2. Let $V_n \sim \chi^2_n$ and $T_n \sim t_n$ for $n \in \{1, 2, 3, \dots\}$.

(a) Find numbers a_n and b_n such that $a_n(V_n - b_n)$ converges in distribution to $\mathcal{N}(0, 1)$.

(b) Show that $T_n^2/(n + T_n^2)$ has a Beta distribution (without using calculus).

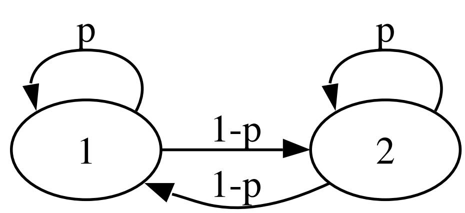
3. Let (X, Y) be Bivariate Normal, with X and Y marginally $\mathcal{N}(0, 1)$ and with correlation ρ between X and Y.

(a) Use MGFs to show that if $\rho = 0$, then X and Y are independent (without quoting the result from class about uncorrelated vectors within a Multivariate Normal). You can use the fact that the MGF $E(e^{sX+tY})$ of a random vector (X, Y) determines its joint distribution (if the MGF exists).

(b) Show that $(X + Y, X - Y)$ is also Bivariate Normal.

(c) Find the joint PDF of $X + Y$ and $X - Y$ (without using calculus).

4. Consider the Markov chain shown below, where $0 < p < 1$ and the labels on the



arrows indicate the probabilities of making those transitions.

(a) Write down the transition matrix Q for this chain.

(b) Find the stationary distribution of the chain.

(c) What is the limit of Q^n as $n \to \infty$?

5. In the Ehrenfest chain, there are two containers with a total of M (distinguishable) particles, and transitions are done by choosing a random particle and moving it from its current container into the other container. Initially (at time $n = 0$), all of the particles are in the second container. Let X_n be the number of particles in the first container at time n (so $X_0 = 0$ and the transition from X_n to X_{n+1} is done as described above). This is a Markov chain with state space $\{0, 1, \ldots, M\}$.

(a) Why is $q_{ii}^{(n)}$ (i.e., the probability of being in state i after n steps, starting from state i) always 0 if n is odd?

(b) Show that $(s_0, s_1, \ldots, s_M)$ with $s_i = \binom{M}{i}(\frac{1}{2})^M$ is the stationary distribution. Why does this, which is a Binomial distribution, seem reasonable intuitively?

Hint: first show that $s_i q_{ij} = s_j q_{ji}$.

6. Daenerys has three dragons: Drogon, Rhaegal, and Viserion. Each dragon independently explores the world in search of tasty morsels. Let X_n, Y_n, Z_n be the locations at time n of Drogon, Rhaegal, Viserion respectively, where time is assumed to be discrete and the number of possible locations is a finite number M. Their paths $X_0, X_1, X_2, \ldots$; $Y_0, Y_1, Y_2, \ldots$; and $Z_0, Z_1, Z_2, \ldots$ are independent Markov chains with the same stationary distribution $\mathbf{s}$. Each dragon starts out at a random location generated according to the stationary distribution.

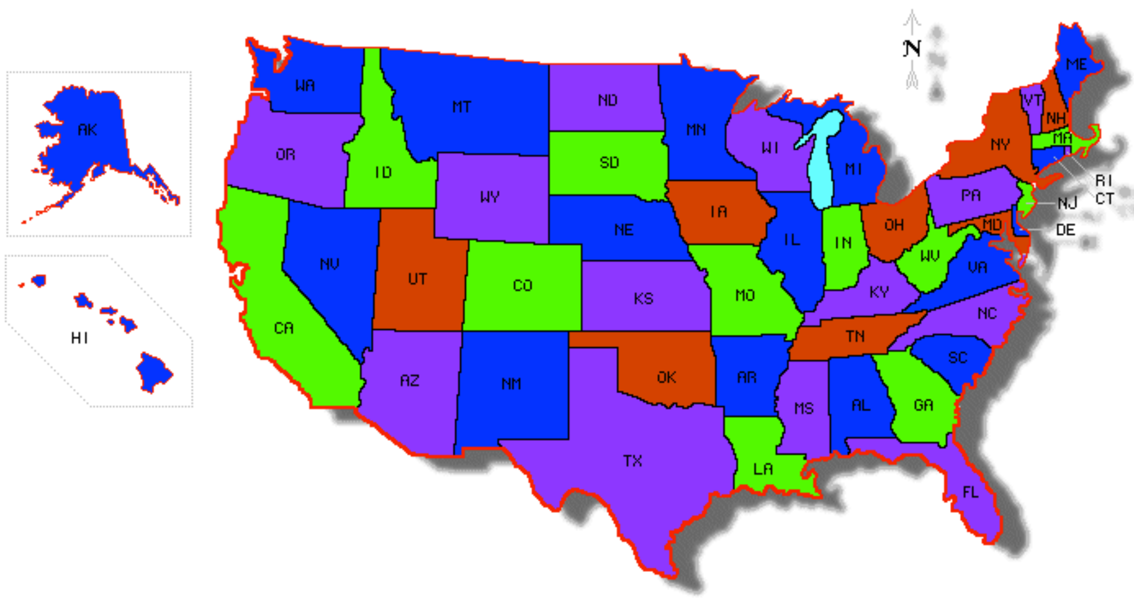
(a) Let state 0 be "home" (so s_0 is the stationary probability of the home state). Find the expected number of times that Drogon is at home, up to time 24, i.e., the expected number of how many of $X_0, X_1, \ldots, X_{24}$ are state 0 (in terms of s_0).

(b) If we want to track all 3 dragons simultaneously, we need to consider the vector of positions, (X_n, Y_n, Z_n). There are M^3 possible values for this vector; assume that each is assigned a number from 1 to M^3, e.g., if $M = 2$ we could encode the states $(0,0,0), (0,0,1), (0,1,0), \ldots, (1,1,1)$ as $1, 2, 3, \ldots, 8$ respectively. Let W_n be the number between 1 and M^3 representing (X_n, Y_n, Z_n). Determine whether $W_0, W_1, \ldots$ is a Markov chain.

(c) Given that all 3 dragons start at home at time 0, find the expected time it will take for all 3 to be at home again at the same time.

7. Let G be a network (also called a *graph*); there are n *nodes*, and for each pair of distinct nodes, there either is or isn't an *edge* joining them. We have a set of k *colors*,

e.g., if $k = 7$ the color set may be {red, orange, yellow, green, blue, indigo, violet}. A *k-coloring* of the network is an assignment of a color to each node, such that two nodes joined by an edge can't be the same color. For example, a 4-coloring of the 50 states in the U.S. is shown below (created using http://monarch.tamu.edu/~maps2/), where the corresponding network has 50 nodes (one for each state), with an edge between two different states if and only if they share a border of nonzero length. Graph coloring is an important topic in computer science, with wide-ranging applications such as coloring a map, scheduling tasks, and in the game of Sudoku.



Suppose that it is possible to k-color G. Form a Markov chain on the space of all k-colorings of G, with transitions as follows: starting with a k-coloring of G, pick a uniformly random node, figure out what the legal colors are for that node, and then repaint that node with a uniformly random legal color (note that this random color may be the same as the current color). Show that this Markov chain is reversible, and find its stationary distribution.

Stat 110 Ultimate Homework Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. Each day, a very volatile stock rises 70% or drops 50% in price, with equal probabilities (with different days independent). Let X_n be the stock price after n days, starting from an initial value of $X_0 = 100$.

(a) Explain why $\log X_n$ is approximately Normal for n large, and state its parameters.

We have $X_n = X_0 Y_1 Y_2 \dots Y_n$, where Y_j is 1.7 if the stock goes up on the jth day and Y_j is 0.5 otherwise (so $Y_j = X_j/X_{j-1}$). Taking logs,

$$\log(X_n) = \log(X_0) + \log(Y_1) + \dots + \log(Y_n).$$

Let $Z_j = \log(Y_j)$. Then the Z_j are i.i.d. with

$$P(Z_n = \log(0.5)) = P(Z_n = \log(1.7)) = 0.5,$$

$$E(Z_n) = \log(0.85)/2 \approx -0.081,$$
$$\mathrm{Var}(Z_n) \approx 0.374.$$

By the CLT, the distribution of $Z_1 + \dots + Z_n$, after standardization, converges to a Normal distribution. Thus, $\log(X_n)$ is approximately Normal for large n, with

$$E(\log(X_n)) = \log(X_0) + nE(Z_1) = \log(100) + \log(0.85)n/2 \approx \log(100) - 0.081n,$$

$$\mathrm{Var}(\log(X_n)) = n\mathrm{Var}(Z_n) \approx 0.374n.$$

Alternatively, we can write $X_n = X_0(0.5)^{n-U_n}(1.7)^{U_n}$ where $U_n \sim \mathrm{Bin}(n, \frac{1}{2})$ is the number of times the stock rises in the first n days. This gives

$$\log(X_n) = \log(X_0) - n\log(2) + U_n \log(3.4),$$

which is just a location and scale transformation of U_n. By the CLT, U_n is approximately $\mathcal{N}(\frac{n}{2}, \frac{n}{4})$ for large n, so $\log(X_n)$ is approximately Normal with the parameters as above for large n.

(b) What happens to $E(X_n)$ as $n \to \infty$?

We have $E(X_1) = (170 + 50)/2 = 110$. Similarly,

$$E(X_{n+1}|X_n) = \frac{1}{2}(1.7X_n) + \frac{1}{2}(0.5X_n) = 1.1X_n,$$

so

$$E(X_{n+1}) = E(E(X_{n+1}|X_n)) = 1.1E(X_n).$$

Thus, $E(X_n) = 1.1^n E(X_0)$ goes to ∞ as $n \to \infty$.

(c) Use the Law of Large Numbers to find out what happens to X_n as $n \to \infty$.

Hint: let U_n be the number of days the stock rises up until time n.

Let $U_n \sim \text{Bin}(n, \frac{1}{2})$ be the number of times the stock rises in the first n days. Note that even though $E(X_n) \to \infty$, if the stock goes up 70% one day and then drops 50% the next day, then overall it has dropped 15% since $1.7 \cdot 0.5 = 0.85$. So X_n will be very small (for n large) if about half the time the stock rose 70% and about half the time the stock dropped 50%; and the Law of Large Numbers ensures that this *will* be the case! We have

$$X_n = X_0(0.5)^{n-U_n}(1.7)^{U_n} = X_0\left(\frac{(3.4)^{U_n/n}}{2}\right)^n,$$

where we simplified in terms of U_n/n so that we can apply the Law of Large Numbers, which here gives that $U_n/n \to 0.5$ with probability 1. But then $(3.4)^{U_n/n} \to \sqrt{3.4} < 2$ with probability 1, so $X_n \to 0$ with probability 1.

Paradoxically, $EX_n \to \infty$ but $X_n \to 0$ with probability 1. With very high probability, the stock price will get very close to 0 since gaining 70% half the time while losing 50% the other half of the time is very bad; but the expected price goes to ∞ since on average each day the price is up 10% from the previous day. To gain more intuition on this, consider a more extreme example, where a gambler starts with \$100 and each day either quadruples his or her money or loses the entire fortune, with equal probabilities. Then on average the gambler's wealth doubles each day, which sounds good for that gambler until one notices that eventually there will be a day when the gambler goes broke. So the actual fortune goes to 0 with probability 1, whereas the expected value goes to infinity due to tiny probabilities of getting extremely large amounts of money (as in the St. Petersburg Paradox).

2. Let $V_n \sim \chi^2_n$ and $T_n \sim t_n$ for $n \in \{1, 2, 3, \dots\}$.

(a) Find numbers a_n and b_n such that $a_n(V_n - b_n)$ converges in distribution to $\mathcal{N}(0, 1)$.

By definition of χ^2_n, we can take $V_n = Z_1^2 + \cdots + Z_n^2$, where $Z_j \sim \mathcal{N}(0, 1)$ independently. We have $E(Z_1^2) = 1$ and $E(Z_1^4) = 3$, so $\text{Var}(Z_1^2) = 2$. By the CLT, if we standardize V_n it will go to $\mathcal{N}(0, 1)$:

$$\frac{Z_1^2 + \cdots + Z_n^2 - n}{\sqrt{2n}} \to \mathcal{N}(0, 1) \text{ in distribution.}$$

So we can take $a_n = \frac{1}{\sqrt{2n}}$, $b_n = n$.

(b) Show that $T_n^2/(n+T_n^2)$ has a Beta distribution (without using calculus).

We can take $T_n = Z_0/\sqrt{V_n/n}$, with $Z_0 \sim \mathcal{N}(0,1)$ independent of V_n. Then we have $T_n^2/(n+T_n^2) = Z_0^2/(Z_0^2+V_n)$, with $Z_0^2 \sim \text{Gamma}(1/2,1/2), V_n \sim \text{Gamma}(n/2,1/2)$. By the bank-post office story, $Z_0^2/(Z_0^2+V_n) \sim \text{Beta}(1/2,n/2)$.

3. Let (X,Y) be Bivariate Normal, with X and Y marginally $\mathcal{N}(0,1)$ and with correlation ρ between X and Y.

(a) Use MGFs to show that if $\rho = 0$, then X and Y are independent (without quoting the result from class about uncorrelated vectors within a Multivariate Normal). You can use the fact that the MGF $E(e^{sX+tY})$ of a random vector (X,Y) determines its joint distribution (if the MGF exists).

Suppose that $\text{Cov}(X,Y) = 0$. Then the MGF of (X,Y) is

$$E(e^{sX+tY}) = e^{E(sX+tY)+\frac{1}{2}\text{Var}(sX+tY)} = e^{\frac{1}{2}(s^2+t^2)} = e^{\frac{1}{2}s^2}e^{\frac{1}{2}t^2}.$$

For Z_1, Z_2 i.i.d. $\mathcal{N}(0,1)$, the joint MGF is

$$E(e^{sZ_1+tZ_2}) = E(e^{sZ_1})E(e^{tZ_2}) = e^{\frac{1}{2}s^2}e^{\frac{1}{2}t^2},$$

the same as the above! Thus, X and Y are i.i.d. $\mathcal{N}(0,1)$.

(b) Show that $(X+Y, X-Y)$ is also Bivariate Normal.

The linear combination $s(X+Y)+t(X-Y) = (s+t)X+(s-t)Y$ is also a linear combination of X and Y, so it is Normal, which shows that $(X+Y, X-Y)$ is MVN.

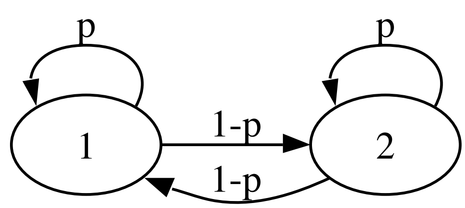
(c) Find the joint PDF of $X+Y$ and $X-Y$ (without using calculus).

Since $X+Y$ and $X-Y$ are uncorrelated (as $\text{Cov}(X+Y, X-Y) = \text{Var}(X) - \text{Var}(Y) = 0$) and $(X+Y, X-Y)$ is MVN, they are independent. Marginally, $X+Y \sim \mathcal{N}(0, 2+2\rho)$ and $X-Y \sim \mathcal{N}(0, 2-2\rho)$. Thus, the joint PDF is

$$f(s,t) = \frac{1}{4\pi\sqrt{1-\rho^2}}e^{-\frac{1}{4}(s^2/(1+\rho)+t^2/(1-\rho))}.$$

4. Consider the Markov chain shown below, where $0 < p < 1$ and the labels on the arrows indicate the probabilities of making those transitions.

(a) Write down the transition matrix Q for this chain.



The transition matrix is

$$Q = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

(b) Find the stationary distribution of the chain.

Because Q is symmetric, the stationary distribution for the chain is the uniform distribution $(1/2, 1/2)$.

(c) What is the limit of Q^n as $n \to \infty$?

The limit of Q^n as $n \to \infty$ is the matrix with the limit distribution $(1/2, 1/2)$ as each row, i.e., $\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$.

5. In the Ehrenfest chain, there are two containers with a total of M (distinguishable) particles, and transitions are done by choosing a random particle and moving it from its current container into the other container. Initially (at time $n = 0$), all of the particles are in the second container. Let X_n be the number of particles in the first container at time n (so $X_0 = 0$ and the transition from X_n to X_{n+1} is done as described above). This is a Markov chain with state space $\{0, 1, \ldots, M\}$.

(a) Why is $q_{ii}^{(n)}$ (i.e., the probability of being in state i after n steps, starting from state i) always 0 if n is odd?

The quantity $q_{ii}^{(n)}$ is the probability that, given that the first container has i particles, it will again have i particles n steps later. We want to prove that $q_{ii}^{(n)} = 0$ for all odd n. Note that the parity of X_n (whether it is even or odd) changes after each step of the chain. So after an odd number of steps, the parity is different from the initial parity, which shows that $q_{ii}^{(n)} = 0$.

(b) Show that $(s_0, s_1, \ldots, s_M)$ with $s_i = \binom{M}{i}(\frac{1}{2})^M$ is the stationary distribution. Why does this, which is a Binomial distribution, seem reasonable intuitively?

Hint: first show that $s_i q_{ij} = s_j q_{ji}$.

The Binomial distribution makes sense here since after running the Markov chain for a long time, each particle is about equally likely to be in either container, approximately independently. To show that the stationary distribution is Binomial$(M, 1/2)$, we will check the reversibility condition $s_i q_{ij} = s_j q_{ji}$.

Let $s_i = \binom{M}{i}(\frac{1}{2})^M$, and check that $s_i q_{ij} = s_j q_{ji}$. If $j = i+1$ (for $i < M$), then

$$s_i q_{ij} = \binom{M}{i}(\frac{1}{2})^M \frac{M-i}{M} = \frac{M!}{(M-i)!i!}(\frac{1}{2})^M \frac{M-i}{M} = \binom{M-1}{i}(\frac{1}{2})^M,$$

$$s_j q_{ji} = \binom{M}{j}(\frac{1}{2})^M \frac{j}{M} = \frac{M!}{(M-j)!j!}(\frac{1}{2})^M \frac{j}{M} = \binom{M-1}{j-1}(\frac{1}{2})^M = s_i q_{ij}.$$

Similarly, if $j = i-1$ (for $i > 0$), then $s_i q_{ij} = s_j q_{ji}$. For all other values of i, j, $q_{ij} = q_{ji} = 0$. Therefore, $\mathbf{s}$ is stationary.

6. Daenerys has three dragons: Drogon, Rhaegal, and Viserion. Each dragon independently explores the world in search of tasty morsels. Let X_n, Y_n, Z_n be the locations at time n of Drogon, Rhaegal, Viserion respectively, where time is assumed to be discrete and the number of possible locations is a finite number M. Their paths $X_0, X_1, X_2, \dots$; $Y_0, Y_1, Y_2, \dots$; and $Z_0, Z_1, Z_2, \dots$ are independent Markov chains with the same stationary distribution $\mathbf{s}$. Each dragon starts out at a random location generated according to the stationary distribution.

(a) Let state 0 be "home" (so s_0 is the stationary probability of the home state). Find the expected number of times that Drogon is at home, up to time 24, i.e., the expected number of how many of $X_0, X_1, \dots, X_{24}$ are state 0 (in terms of s_0).

By definition of stationarity, at each time Drogon has probability s_0 of being at home. By linearity, the desired expected value is $25 s_0$.

(b) If we want to track all 3 dragons simultaneously, we need to consider the vector of positions, (X_n, Y_n, Z_n). There are M^3 possible values for this vector; assume that each is assigned a number from 1 to M^3, e.g., if $M = 2$ we could encode the states $(0,0,0), (0,0,1), (0,1,0), \dots, (1,1,1)$ as $1, 2, 3, \dots, 8$ respectively. Let W_n be the number between 1 and M^3 representing (X_n, Y_n, Z_n). Determine whether $W_0, W_1, \dots$ is a Markov chain.

Yes, $W_0, W_1, \dots$ is a Markov chain, since given the entire past history of the X, Y, and Z chains, only the most recent information about the whereabouts of the dragons should be used in predicting their vector of locations. To show this algebraically, let

A_n be the event $\{X_0 = x_0, \ldots, X_n = x_n\}$, B_n be the event $\{Y_0 = y_0, \ldots, Y_n = y_n\}$, C_n be the event $\{Z_0 = z_0, \ldots, Z_n = z_n\}$, and $D_n = A_n \cap B_n \cap C_n$. Then

$$P(X_{n+1} = x, Y_{n+1} = y, Z_{n+1} = z | D_n)$$

$$= P(X_{n+1} = x | D_n) P(Y_{n+1} = y | X_{n+1} = x, D_n) P(Z_{n+1} = z | X_{n+1} = x, Y_{n+1} = y, D_n)$$

$$= P(X_{n+1} = x | A_n) P(Y_{n+1} = y | B_n) P(Z_{n+1} = z | C_n)$$

$$= P(X_{n+1} = x | X_n = x_n) P(Y_{n+1} = y | Y_n = y_n) P(Z_{n+1} = z | Z_n = z_n).$$

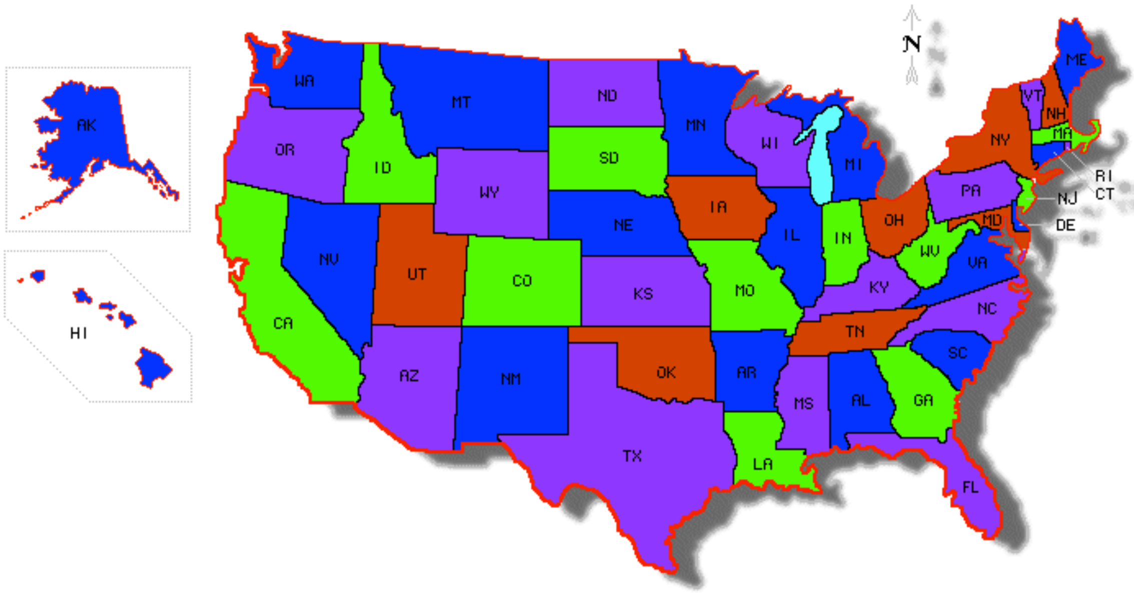
(c) Given that all 3 dragons start at home at time 0, find the expected time it will take for all 3 to be at home again at the same time.

The stationary probability for the W-chain of the state with Drogon, Rhaegal, Viserion being at locations x, y, z is $s_x s_y s_z$, since if (X_n, Y_n, Z_n) is drawn from this distribution, then marginally each dragon's location is distributed according to its stationary distribution, so $P(X_{n+1} = x, Y_{n+1} = y, Z_{n+1} = z) = P(X_{n+1} = x)P(Y_{n+1} = y)P(Z_{n+1} = z) = s_x s_y s_z$. So the expected time for all 3 dragons to be home at the same time, given that they all start at home, is $1/s_0^3$.

7. Let G be a network (also called a *graph*); there are n *nodes*, and for each pair of distinct nodes, there either is or isn't an *edge* joining them. We have a set of k *colors*, e.g., if $k = 7$ the color set may be {red, orange, yellow, green, blue, indigo, violet}. A *k-coloring* of the network is an assignment of a color to each node, such that two nodes joined by an edge can't be the same color. For example, a 4-coloring of the 50 states in the U.S. is shown below (created using http://monarch.tamu.edu/∼maps2/), where the corresponding network has 50 nodes (one for each state), with an edge between two different states if and only if they share a border of nonzero length. Graph coloring is an important topic in computer science, with wide-ranging applications such as coloring a map, scheduling tasks, and in the game of Sudoku.

Suppose that it is possible to k-color G. Form a Markov chain on the space of all k-colorings of G, with transitions as follows: starting with a k-coloring of G, pick a uniformly random node, figure out what the legal colors are for that node, and then repaint that node with a uniformly random legal color (note that this random color may be the same as the current color). Show that this Markov chain is reversible, and find its stationary distribution.

Let C be the set of all k-colorings of G, and let q_{ij} be the transition probability of going from i to j for any k-colorings i and j in C. We will show that $q_{ij} = q_{ji}$, which implies that the stationary distribution is uniform on C.



For any k-coloring i and node v, let $L(i, v)$ be the number of legal colorings for node v, keeping the colors of all other nodes the same as they are in i. If k-colorings i and j differ at more than one node, then $q_{ij} = 0 = q_{ji}$. If $i = j$, then obviously $q_{ij} = q_{ji}$. If i and j differ at exactly one node v, then $L(i, v) = L(j, v)$, so

$$q_{ij} = \frac{1}{n}\frac{1}{L(i,v)} = \frac{1}{n}\frac{1}{L(j,v)} = q_{ji}.$$

So

$$\frac{1}{c}q_{ij} = \frac{1}{c}q_{ji}$$

for all i and j in C, where c is the size of C. This shows that the chain is reversible, with the uniform distribution on C as stationary distribution.

This Markov chain is an example of what is known as a *Gibbs sampler*. Together, the Gibbs sampler and the Metropolis algorithm described in SP 11 have revolutionized scientific computing, as mainstays of Markov chain Monte Carlo.