6.041/6.431 Spring 2009 Final Exam
Thursday, May 21, 1:30 - 4:30 PM.

Name: ____________________

Recitation Instructor: ____________________

Question	Part	Score	Out of
0			2
1	all		18
2	all		24
3	a		4
	b		4
	c		4
4	a		6
	b		6
	c		6
5	a		6
	b		6
6	a		4
	b		4
	c		4
	d		5
	e		5
7	a		6
	b		6
Total			120

- Write your solutions in this quiz packet, only solutions in the quiz packet will be graded.
- You are allowed three two-sided 8.5 by 11 formula sheet plus a calculator.
- You have 180 minutes to complete the quiz.
- Be neat! You will not get credit if we can't read it.

Problem 1: True or False (2pts. each, 18 pts. total)

No partial credit will be given for individual questions in this part of the quiz.

a. Let $\{X_n\}$ be a sequence of i.i.d random variables taking values in the interval $[0, 0.5]$. Consider the following statements:

(A) If $\mathbf{E}[X_n^2]$ converges to 0 as $n \to \infty$ then X_n converges to 0 in probability.

(B) If all X_n have $\mathbf{E}[X_n] = 0.2$ and $\text{var}(X_n)$ converges to 0 as $n \to \infty$ then X_n converges to 0.2 in probability.

(C) The sequence of random variables Z_n, defined by $Z_n = X_1 \cdot X_2 \cdots X_n$, converges to 0 in probability as $n \to \infty$.

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: True	**B**: True	**C**: True

Solution:

(A) True. The fact that $\lim_{n\to\infty} \mathbf{E}[X_n^2] = 0$ implies $\lim_{n\to\infty} \mathbf{E}[X_n] = 0$ and $\lim_{n\to\infty} \text{var}(X_n) = 0$. Hence, one has

$$\begin{aligned}\mathbf{P}(|X_n - 0| \geq \epsilon) &\leq \mathbf{P}(|X_n - \mathbf{E}[X_n]| \geq \epsilon/2) + \mathbf{P}(|\mathbf{E}[X_n] - 0| \geq \epsilon/2) \\ &\leq \frac{\text{var}(X_n)}{(\epsilon/2)^2} + \mathbf{P}(|\mathbf{E}[X_n] - 0| \geq \epsilon/2) \to 0,\end{aligned}$$

where we have applied Chebyshev inequality.

(B) True. Applying Chebyshev inequality gives

$$\mathbf{P}(|X_n - \mathbf{E}[X_n]| \geq \epsilon) \leq \frac{\text{var}(X_n)}{\epsilon^2} \to 0.$$

Hence X_n converges to $\mathbf{E}[X_n] = 0.2$ in probability.

(C) True. For all $\epsilon > 0$, since $Z_n \leq (1/2)^n \Rightarrow \mathbf{P}(|Z_n - 0| \geq \epsilon) = 0$ for $n > -\log \epsilon / \log 2$.

b. Let X_i $(i = 1, 2, \dots)$ be i.i.d. random variables with mean 0 and variance 2; Y_i $(i = 1, 2, \dots)$ be i.i.d. random variables with mean 2. Assume that all variables X_i, Y_j are independent. Consider the following statements:

(A) $\frac{X_1 + \cdots + X_n}{n}$ converges to 0 in probability as $n \to \infty$.

(B) $\frac{X_1^2 + \cdots + X_n^2}{n}$ converges to 2 in probability as $n \to \infty$.

(C) $\frac{X_1 Y_1 + \cdots + X_n Y_n}{n}$ converges to 0 in probability as $n \to \infty$.

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: True	**B**: True	**C**: True

Solution:

(A) True. Note that $\mathbf{E}[\frac{X_1+\cdots+X_n}{n}] = 0$ and $\text{var}(\frac{X_1+\cdots+X_n}{n}) = \frac{n\cdot 2}{n^2} = \frac{2}{n}$. One can see $\frac{X_1+\cdots+X_n}{n}$ converges to 0 in probability.

(B) True. Let $Z_i = X_i^2$ and $\mathbf{E}[Z_i] = 2$. Note Z_i are i.i.d. since X_i are i.i.d., and hence one has that $\frac{Z_1+\cdots+Z_n}{n}$ converges to $\mathbf{E}[Z_i] = 2$ in probability by the WLLN.

(C) True. Let $W_i = X_iY_i$ and $\mathbf{E}[W_i] = \mathbf{E}[X_i]\mathbf{E}[Y_i] = 0$. Note W_i are i.i.d. since X_i and Y_i are respectively i.i.d., and hence one has that $\frac{W_1+\cdots+W_n}{n}$ converges to $\mathbf{E}[W_i] = 0$ in probability by the WLLN.

c. We have i.i.d. random variables $X_1 \dots X_n$ with an unknown distribution, and with $\mu = \mathbf{E}[X_i]$. We define $M_n = (X_1 + \ldots + X_n)/n$. Consider the following statements:

(A) M_n is a maximum-likelihood estimator for μ, irrespective of the distribution of the X_i's.

(B) M_n is a consistent estimator for μ, irrespective of the distribution of the X_i's.

(C) M_n is an asymptotically unbiased estimator for μ, irrespective of the distribution of the X_i's.

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: False	**B**: True	**C**: True

Solution:

(A) False. Consider X_i follow a uniform distribution $U[\mu - \frac{1}{2}, \mu + \frac{1}{2}]$. The ML estimator for μ is any value between $\max(X_1, \cdots, X_n) - \frac{1}{2}$ and $\min(X_1, \cdots, X_n) + \frac{1}{2}$, instead of M_n.

(B) True. By the WLLN, M_n converges to μ in probability and hence it is a consistent estimator.

(C) True. Since $\mathbf{E}[M_n] = \mathbf{E}[X_i] = \mu$, M_n is unbiased estimator for μ and hence asymptotically unbiased.

Problem 2: Multiple Choice (4 pts. each, 24 pts. total)

Clearly circle the appropriate choice. **No partial credit** will be given for individual questions in this part of the quiz.

a. Earthquakes in Sumatra occur according to a Poisson process of rate $\lambda = 2/\text{year}$. Conditioned on the event that exactly two earthquakes take place in a year, what is the probability that both earthquakes occur in the first three months of the year? (for simplicity, assume all months have 30 days, and each year has 12 months, i.e., 360 days).

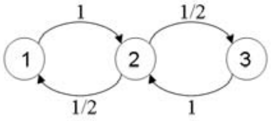
   (i) 1/12
   (ii) $\boxed{1/16}$
   (iii) 64/225
   (iv) $4e^{-4}$
   (v) There is not enough information to determine the required probability.
   (vi) None of the above.

   Solution: Consider the interval of a year be $[0, 1]$.

$$\begin{aligned}\mathbf{P}\left(2 \text{ in } [0, \frac{1}{4}) \mid 2 \text{ in } [0,1]\right) &= \frac{\mathbf{P}\left(2 \text{ in } [0, \frac{1}{4}), 0 \text{ in } [\frac{1}{4}, 1]\right)}{\mathbf{P}(2 \text{ in } [0,1]))} \\ &= \frac{\frac{(\lambda\cdot 1/4)^2}{2!}e^{-\lambda\cdot 1/4}\cdot\frac{(\lambda\cdot 3/4)^0}{0!}e^{-\lambda\cdot 3/4}}{\frac{\lambda^2}{2!}e^{-\lambda}} \\ &= \frac{1}{16}\end{aligned}$$

   (alternative explanation) Given that exactly two earthquakes happened in 12 months, each earthquake is equally likely to happen in any month of the 12, the probability that it happens in the first 3 months is $3/12 = 1/4$. The probability that both happen in the first 3 months is $(1/4)^2$.

b. Consider a continuous-time Markov chain with three states $i \in \{1, 2, 3\}$, with dwelling time in each visit to state i being an exponential random variable with parameter $\nu_i = i$, and transition probabilities p_{ij} defined by the graph



   What is the long-term expected fraction of time spent in state 2?

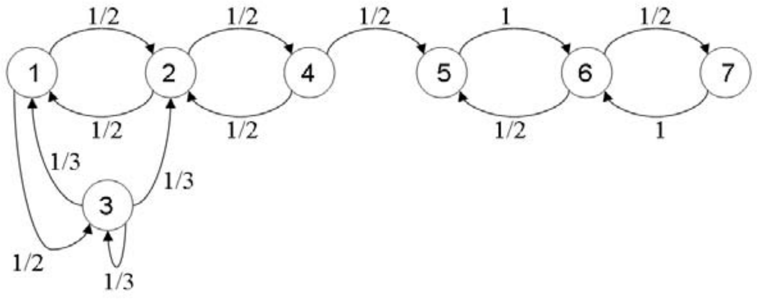
   (i) 1/2
   (ii) 1/4
   (iii) 2/5

(iv) $\boxed{3/7}$

(v) None of the above.

Solution: First, we calculate the $q_{ij} = \nu_i p_{ij}$, i.e., $q_{12} = q_{21} = q_{23} = 1$ and $q_{32} = 3$. The balance and normalization equations of this birth-death markov chain can be expressed as, $\pi_1 = \pi_2, \pi_2 = 3\pi_3$ and $\pi_1 + \pi_2 + \pi_3 = 1$, yielding $\pi_2 = 3/7$.

c. Consider the following Markov chain:



Starting in state 3, what is the steady-state probability of being in state 1?

- (i) 1/3
- (ii) 1/4
- (iii) 1
- (iv) $\boxed{0}$
- (v) None of the above.

Solution: State 1 is transient.

d. Random variables X and Y are such that the pair (X, Y) is uniformly distributed over the trapezoid A with corners $(0, 0)$, $(1, 2)$, $(3, 2)$, and $(4, 0)$ shown in Fig. 1:

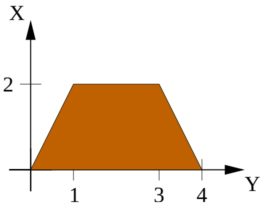


Figure 1: $f_{X,Y}(x, y)$ is constant over the shaded area, zero otherwise.

i.e.

$$f_{X,Y}(x,y) = \begin{cases} c\,, & (x,y) \in A \\ 0\,, & \text{else}\,. \end{cases}$$

We observe Y and use it to estimate X. Let $\hat{X}$ be the least mean squared error estimator of X given Y. What is the value of $\text{var}(\hat{X} - X | Y = 1)$?

- (i) 1/6
- (ii) 3/2
- (iii) $\boxed{1/3}$
- (iv) The information is not sufficient to compute this value.

(v) None of the above.

Solution: $f_{X|Y=1}(x)$ is uniform on $[0,2]$ therefore $\hat{X} = \mathbf{E}[X|Y=1] = 1$ and $\text{var}(\hat{X} - X|Y = 1) = \text{var}(X|Y=1) = (2-0)^2/12 = 1/3$.

e. $X_1 \ldots X_n$ are i.i.d. normal random variables with mean value μ and variance v. Both μ and v are unknown. We define $M_n = (X_1 + \ldots + X_n)/n$ and

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$$

We also define $\Phi(x)$ to be the CDF for the standard normal distribution, and $\Psi_{n-1}(x)$ to be the CDF for the t-distribution with $n-1$ degrees of freedom. Which of the following choices gives an exact 99% confidence interval for μ for all $n > 1$?

(i) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Phi(\delta) = 0.99$.

(ii) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Phi(\delta) = 0.995$.

(iii) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Psi_{n-1}(\delta) = 0.99$.

(iv) $\boxed{[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}] \text{ where } \delta \text{ is chosen to give } \Psi_{n-1}(\delta) = 0.995.}$

(v) None of the above.

Solution: See Lecture 23, slides 10-12.

f. We have i.i.d. random variables X_1, X_2 which have an exponential distribution with unknown parameter θ. Under hypothesis H_0, $\theta = 1$. Under hypothesis H_1, $\theta = 2$. Under a likelihood-ratio test, the rejection region takes which of the following forms?

(i) $R = \{(x_1, x_2) : x_1 + x_2 > \xi\}$ for some value ξ.

(ii) $\boxed{R = \{(x_1, x_2) : x_1 + x_2 < \xi\} \text{ for some value } \xi.}$

(iii) $R = \{(x_1, x_2) : e^{x_1} + e^{x_2} > \xi\}$ for some value ξ.

(iv) $R = \{(x_1, x_2) : e^{x_1} + e^{x_2} < \xi\}$ for some value ξ.

(v) None of the above.

Solution: We defined $R = \{x = (x_1, x_2) | L(x) > c\}$ where

$$L(x) = \frac{f_X(x; H_1)}{f_X(x; H_0)} = \frac{\theta_1 e^{-\theta_1 x_1}\theta_1 e^{-\theta_1 x_2}}{\theta_0 e^{-\theta_0 x_1}\theta_0 e^{-\theta_0 x_2}} = \frac{\theta_1^2}{\theta_0^2} e^{(\theta_0 - \theta_1)(x_1 + x_2)} = 4e^{-(x_1+x_2)}.$$

So $R = \{(x_1, x_2) | x_1 + x_2 < -\log(c/4)\}$

Problem 3 (12 pts. total)

Aliens of two races (blue and green) are arriving on Earth independently according to Poisson process distributions with parameters λ_b and λ_g respectively. The Alien Arrival Registration Service Authority (AARSA) will begin registering alien arrivals soon.

Let T_1 denote the time AARSA will function until it registers its first alien. Let G be the event that the first alien to be registered is a green one. Let T_2 be the time AARSA will function until at least one alien of both races is registered.

(a) (4 points.) Express $\mu_1 = \mathbf{E}[T_1]$ in terms of λ_g and λ_b. **Show your work.**

Answer: $\boxed{\mu_1 = \mathbf{E}[T_1] = \frac{1}{\lambda_g+\lambda_b}}$

Solution: We consider the process of arrivals of both types of Aliens. This is a merged Poisson process with arrival rate $\lambda_g+\lambda_b$. T_1 is the time until the first arrival, and therefore is exponentially distributed with parameter $\lambda_g+\lambda_b$. Therefore $\mu_1 = \mathbf{E}[T_1] = \frac{1}{\lambda_g+\lambda_b}$.

One can also go about this using derived distributions, since $T_1 = \min(T_1^g, T_1^b)$ where T_1^g and T_1^b are the first arrival times of green and blue Aliens respectively (i.e., T_1^g and T_1^b are exponentially distributed with parameters λ_g and λ_b, respectively.)

(b) (4 points.) Express $p = \mathbf{P}(G)$ in terms of λ_g and λ_b. **Show your work.**

Answer: $\boxed{\mathbf{P}(G) = \frac{\lambda_g}{\lambda_g+\lambda_b}}$

Solution: We consider the same merged Poisson process as before, with arrival rate $\lambda_g+\lambda_b$. Any particular arrival of the merged process has probability $\frac{\lambda_g}{\lambda_g+\lambda_b}$ of corresponding to a green Alien and probability $\frac{\lambda_b}{\lambda_g+\lambda_b}$ of corresponding to a blue Alien. The question asks for $\mathbf{P}(G) = \frac{\lambda_g}{\lambda_g+\lambda_b}$.

(c) (4 points.) Express $\mu_2 = \mathbf{E}[T_2]$ in terms of λ_g and λ_b.
Show your work.

$$\boxed{\textbf{Answer: } \frac{1}{\lambda_g+\lambda_b} + \frac{\lambda_g}{\lambda_g+\lambda_b}\left(\frac{1}{\lambda_b}\right) + \frac{\lambda_b}{\lambda_g+\lambda_b}\left(\frac{1}{\lambda_g}\right)}$$

Solution: The time T_2 until at least one green and one red Aliens have arrived can be expressed as $T_2 = \max(T_1^g, T_1^b)$, where T_1^g and T_1^b are the first arrival times of green and blue Aliens respectively (i.e., T_1^g and T_1^b are exponentially distributed with parameters λ_g and λ_b, respectively.)

The expected time till the 1st Alien arrives was calculated in (a), $\mu_1 = \mathbf{E}[T_1] = \frac{1}{\lambda_g+\lambda_b}$. To compute the remaining time we simply condition on the 1st Alien being green(e.g. event G) or blue(event G^c), and use the memoryless property of Poisson, i.e.,

$$\begin{aligned}
\mathbf{E}[T_2] &= \mathbf{E}[T_1] + \mathbf{P}(G)\mathbf{E}[\text{Time until first Blue arrives}|G] + \mathbf{P}(G^c)\mathbf{E}[\text{Time until first Green arrives}|G^c] \\
&= \mathbf{E}[T_1] + \mathbf{P}(G)\mathbf{E}[T_2^b] + (1-\mathbf{P}(G))\mathbf{E}[T_2^g] \\
&= \frac{1}{\lambda_g+\lambda_b} + \frac{\lambda_g}{\lambda_g+\lambda_b}\left(\frac{1}{\lambda_b}\right) + \frac{\lambda_b}{\lambda_g+\lambda_b}\left(\frac{1}{\lambda_g}\right)
\end{aligned}$$

Problem 4 (18 pts. total)

Researcher Jill is interested in studying employment in technology firms in Dilicon Valley. She denotes by X_i the number of employees in technology firm i and assumes that X_i are independent and identically distributed with mean p. To estimate p, Jill randomly interviews n technology firms and observes the number of employees in these firms.

(a) (6 points.) Jill uses

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

as an estimator for p. Find the limit of $\mathbf{P}(M_n \leq x)$ as $n \to \infty$ for $x < p$. Find the limit of $\mathbf{P}(M_n \leq x)$ as $n \to \infty$ for $x > p$. **Show your work.**

Solution: Since X_i is i.i.d., M_n converges to p in probability, i.e., $\lim_{n\to\infty} \mathbf{P}(|M_n - p| > \epsilon) = 0$, implying $\lim_{n\to\infty} \mathbf{P}(M_n < p - \epsilon) = 0$ and $\lim_{n\to\infty} \mathbf{P}(M_n > p + \epsilon) = 0$, for all $\epsilon > 0$. Hence

$$\boxed{\lim_{n\to\infty} \mathbf{P}(M_n \leq x) = \begin{cases} 0, & x < p\,, \\ 1, & x > p\,. \end{cases}}$$

(b) (6 points.) Find the smallest n, the number of technology firms Jill must sample, for which the Chebyshev inequality yields a guarantee

$$\mathbf{P}(|M_n - p| \geq 0.5) \leq 0.05.$$

Assume that $\text{var}\,(X_i) = v$ for some constant v. State your solution as a function of v. **Show your work.**

Solution: Since M_n converges to p in probability and $\text{var}(M_n) = \frac{n}{n^2} \cdot \text{var}(X_i) = v/n$, Chebyshev inequality gives

$$P(|M_n - p| \geq 0.5) \leq \frac{\text{var}(M_n)}{0.5^2} = \frac{v}{n \cdot 0.5^2} = 0.05$$

$\Rightarrow \boxed{n = 80v.}$

(c) (6 points.) Assume now that the researcher samples $n = 5000$ firms. Find an approximate value for the probability

$$\mathbf{P}(|M_{5000} - p| \geq 0.5)$$

using the Central Limit Theorem. Assume again that $\text{var}(X_i) = v$ for some constant v. Give your answer in terms of v, and the standard normal CDF Φ. **Show your work.**

Solution: By CLT, we can approximate by a standard normal distribution

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{nv}}$$

when n is large, and hence,

$$\mathbf{P}\left(|M_{5000} - p| \geq 0.5\right) = P\left(\left|\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{nv}}\right| \geq \frac{0.5\sqrt{n}}{\sqrt{v}}\right) = \boxed{2 - 2\Phi\left(\frac{0.5\sqrt{n}}{\sqrt{v}}\right)},$$

where $n = 5000$.

Problem 5 (12 pts. total)

The RandomView window factory produces window panes. After manufacturing, 1000 panes were loaded onto a truck. The weight W_i of the i-th pane (in pounds) on the truck is modeled as a random variable, with the assumption that the W_i's are independent and identically distributed.

(a) (6 points.) Assume that the measured weight of the load on the truck was 2340 pounds, and that $\text{var}(W_i) \leq 4$. Find an approximate 95 percent confidence interval for $\mu = \mathbf{E}[W_i]$, using the Central Limit Theorem (you may use the standard normal table which was handed out with this quiz). **Show your work.**

Answer: $[2.216,\ 2.464]$

Solution: The sample mean estimator $\hat{\Theta}_n = \frac{W_1+\cdots+W_n}{n}$ in this case is

$$\hat{\Theta}_{1000} = \frac{2340}{1000} = 2.34$$

Using the CDF $\Phi(z)$ of the standard normal available in the normal tables, we have $\Phi(1.96) = 0.975$, so we obtain

$$\mathbf{P}\left(\frac{|\hat{\Theta}_{1000} - \mu|}{\sqrt{\text{var}(W_i)/1000}} \leq 1.96\right) \approx 0.95.$$

Because the variance is less than 4, we have

$$\mathbf{P}(|\hat{\Theta}_{1000} - \mu| \leq 1.96\sqrt{\text{var}(W_i)/1000}) \leq \mathbf{P}(|\hat{\Theta}_{1000} - \mu| \leq 1.96\sqrt{4/1000}),$$

and letting the right-hand side of the above equation ≈ 0.95 gives a 95% confidence, i.e.,

$$\left[\hat{\Theta}_{1000} - 1.96\sqrt{\frac{4}{1000}}, \hat{\Theta}_{1000} + 1.96\sqrt{\frac{4}{1000}}\right] = \left[\hat{\Theta}_{1000} - 0.124, \hat{\Theta}_{1000} + 0.124\right]$$

(b) (6 points.) Now assume instead that the random variables W_i are i.i.d, with an exponential distribution with parameter $\theta > 0$, i.e., a distribution with PDF

$$f_W(w;\theta) = \theta e^{-\theta w}$$

What is the maximum likelihood estimate of θ, given that the truckload has weight 2340 pounds? **Show your work.**

$\boxed{\textbf{Answer:} \quad \hat{\Theta}_{1000}^{mle} = \frac{1000}{2340} = 0.4274}$

Solution: The likelihood function is

$$f_W(w;\theta) = \prod_{i=1}^{n} f_{W_i}(w_i;\theta) = \prod_{i=1}^{n} \theta e^{-\theta w_i},$$

And the log-likelihood function is

$$\log f_W(w;\theta) = n\log\theta - \theta\sum_{i=1}^{n} w_i,$$

The derivative with respect to θ is $\frac{n}{\theta} - \sum_{i=1}^{n} w_i,$, and by setting it to zero, we see that the maximum of $\log f_W(w;\theta)$ over $\theta \geq 0$ is attained at $\hat{\theta}_n = \frac{n}{\sum_{i=1}^{n} w_i}$. The resulting estimator is

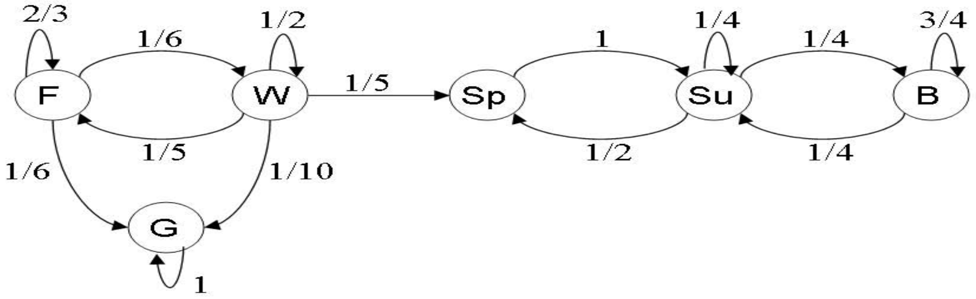
$$\hat{\Theta}_n^{mle} = \frac{n}{\sum_{i=1}^{n} W_i}.$$

In our case,

$$\hat{\Theta}_{1000}^{mle} = \frac{1000}{2340} = 0.4274$$

Problem 6 (21 pts. total)

In Alice's Wonderland, there are six different seasons: Fall (F), Winter (W), Spring (Sp), Summer (Su), Bitter Cold (B), and Golden Sunshine (G). The seasons do not follow any particular order, instead, at the beginning of each day the Head Wizard assigns the season for the day, according to the following Markov chain model:



Thus, for example, if it is Fall one day then there is 1/6 probability that it will be Winter the next day (note that it is possible to have the same season again the next day).

(a) (4 points.) For each state in the above chain, identify whether it is recurrent or transient. **Show your work.**

Solution: F and W are transient states; Sp, Su, B, and G are recurrent states.

(b) (4 points.) If it is Fall on Monday, what is the probability that it will be Summer on Thursday of the same week? **Show your work.**

Solution: There is only one path from F to Su in three days.

$$\begin{aligned}\mathbf{P}(S_4 = \text{Su}|S_1 = \text{F}) &= \mathbf{P}(S_2 = \text{W}|S_1 = \text{F}) \cdot \mathbf{P}(S_3 = \text{Sp}|S_2 = \text{W}) \cdot \mathbf{P}(S_4 = \text{Su}|S_3 = \text{Sp}) \\ &= \frac{1}{6} \cdot \frac{1}{5} \cdot 1 = \boxed{\frac{1}{30}}\end{aligned}$$

(c) (4 points.) If it is Spring today, will the chain converge to steady-state probabilities? If so, compute the steady-state probability for each state. If not, explain why these probabilities do not exist. **Show your work.**

Solution: The Markov chain will stay in the recurrent class $\{\mathrm{Sp}, \mathrm{Su}, \mathrm{B}\}$, and

$$\begin{cases} \pi_{\mathrm{Sp}} \cdot 1 = \pi_{\mathrm{Su}} \cdot \frac{1}{2} \\ \pi_{\mathrm{B}} \cdot \frac{1}{4} = \pi_{\mathrm{Su}} \cdot \frac{1}{4} \\ \pi_{\mathrm{F}} = 0 \\ \pi_{\mathrm{W}} = 0 \\ \pi_{\mathrm{G}} = 0 \\ \pi_{\mathrm{F}} + \pi_{\mathrm{W}} + \pi_{\mathrm{G}} + \pi_{\mathrm{Sp}} + \pi_{\mathrm{Su}} + \pi_{\mathrm{B}} = 1 \end{cases}$$

$\Rightarrow \boxed{\pi_{\mathrm{F}} = 0, \pi_{\mathrm{W}} = 0, \pi_{\mathrm{G}} = 0, \pi_{\mathrm{Sp}} = 1/5, \pi_{\mathrm{Su}} = 2/5, \pi_{\mathrm{B}} = 2/5.}$

(d) (5 points.) If it is Fall today, what is the probability that Bitter Cold will never arrive in the future? **Show your work.**

Solution: Let a_{F} and a_{W} be the probabilities that Bitter Cold will never arrive starting from Fall and Winter, respectively. This is equivalent to the Markov chain ends up in G.

$$\begin{cases} a_{\mathrm{F}} = \frac{2}{3} \cdot a_{\mathrm{F}} + \frac{1}{6} \cdot a_{\mathrm{W}} + \frac{1}{6} \cdot 1 \\ a_{\mathrm{W}} = \frac{1}{5} \cdot a_{\mathrm{F}} + \frac{1}{2} \cdot a_{\mathrm{W}} + \frac{1}{10} \cdot 1 \end{cases}$$

$\Rightarrow \boxed{a_{\mathrm{F}} = 3/4.}$

(e) (5 points.) If it is Fall today, what is the expected number of days till either Summer or Golden Sunshine arrives for the first time? **Show your work.**

Solution: Let μ_{F} and μ_{W} be expected number of days till either Summer or Golden Sunshine arrives for the first time, respectively.

$$\begin{cases} \mu_{\mathrm{F}} = 1 + \frac{2}{3} \cdot \mu_{\mathrm{F}} + \frac{1}{6} \cdot \mu_{\mathrm{W}} + \frac{1}{6} \cdot 0 \\ \mu_{\mathrm{W}} = 1 + \frac{1}{5} \cdot \mu_{\mathrm{F}} + \frac{1}{2} \cdot \mu_{\mathrm{W}} + \frac{1}{5} \cdot 1 \end{cases}$$

$\Rightarrow \boxed{\mu_{\mathrm{F}} = 5.25.}$

Problem 7 (12 pts. total)

A newscast covering the final baseball game between Sed Rox and Y Nakee becomes noisy at the crucial moment when the viewers are informed whether Y Nakee won the game.

Let a be the parameter describing the actual outcome: $a = 1$ if Y Nakee won, $a = -1$ otherwise. There were n viewers listening to the telecast. Let Y_i be the information received by viewer i ($1 \le i \le n$). Under the noisy telecast, $Y_i = a$ with probability p, and $Y_i = -a$ with probability $1 - p$. Assume that the random variables Y_i are independent of each other.

The viewers as a group come up with a joint estimator

$$Z_n = \begin{cases} 1 & \text{if } \sum_{i=1}^n Y_i \ge 0, \\ -1 & \text{otherwise.} \end{cases}$$

(a) (6 points.) Find $\lim_{n\to\infty} \mathbf{P}(Z_n = a)$ assuming that $p > 0.5$ and $a = 1$. **Show your work.**

Solution: Note that

$$\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = \lim_{n\to\infty} \mathbf{P}\left(\sum_{i=1}^n Y_i \ge 0\right) = \lim_{n\to\infty} \mathbf{P}\left(\frac{\sum_{i=1}^n Y_i}{n} \ge 0\right).$$

Since Y_i are i.i.d. with mean $\mathbf{E}[Y_i] = 2p - 1$ and finite variance $\text{var}(Y_i) = 1 - (2p-1)^2$, one has, by Chebyshev inequality, for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\frac{\sum_{i=1}^n Y_i}{n} - (2p-1)\right| \ge \epsilon\right) = 0.$$

Take $\epsilon = p - \frac{1}{2}$, and the above equation implies $\lim_{n\to\infty} \mathbf{P}\left(\frac{\sum_{i=1}^n Y_i}{n} \le (2p-1)/2\right) = 0$. Therefore, $\boxed{\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = 1.}$

(b) (6 points.) Find $\lim_{n\to\infty} \mathbf{P}(Z_n = a)$, assuming that $p = 0.5$ and $a = 1$. **Show your work.**

Solution: Note that

$$\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = \lim_{n\to\infty} \mathbf{P}\left(\sum_{i=1}^n Y_i \ge 0\right) = \lim_{n\to\infty} \mathbf{P}\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}} \ge 0\right).$$

Since Y_i are i.i.d. with $\mathbf{E}[Y_i] = 0$ and $\text{var}(Y_i) = 1$, we can approximate $\frac{\sum_{i=1}^n Y_i}{\sqrt{n}}$ as a standard normal random variable when n goes to infinity. Thus, $\boxed{\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = 1/2.}$

6.041 / 6.431 Probabilistic Systems Analysis and Applied Probability
Fall 2010