332:541 **Stochastic Signals and Systems** **Fall 2000**

Quiz III (Final Exam) Statement and Solutions

You have at least 3 hours to answer the following five questions (two of them short). You are allowed three sides of an 11x8.5 inch piece of paper for notes. GOOD LUCK!

1. *(50 points)* **Willie's Last Stand:** Willie the Wombat is trying to cross a road. Starting from the curb (position 0), it is 4 wombat hops to the opposite curb (position 4). Willie is wearing his Walkman tape player and walks to the beat of the music. Willie can take one hop forward, one hop backward or stay put with each beat of the music.

   Suppose Willie takes one hop forward with probability p when he is in positions 0, 1, 2 or 3 and takes 1 hop backward with probability q when he is in positions 1, 2, 3, or 4. Assume $q + p < 1$.

   (a) *(10 points)* Please carefully sketch the Markov state diagram corresponding to Willie's position. *From LaTex it's easier to list the state transition matrix. If $\pi(n)$ is the probability column vector at beat n, then we have*

$$\pi(n+1) = \begin{bmatrix} 1-p & q & 0 & 0 & 0 \\ p & 1-q-p & q & 0 & 0 \\ 0 & p & 1-q-p & q & 0 \\ 0 & 0 & p & 1-q-p & q \\ 0 & 0 & 0 & p & 1-q \end{bmatrix} \pi(n)$$

   (b) *(10 points)* Suppose Willie starts in position 0 at time $t = 0$ and that positions 1, 2 and 3 are in the road. What is the probability that after 3 beats, Willie is in the road.

   Easiest way is to realize that Willie can only get as far as position 3 in three hops, so that the only way he's not in the street is if he's in 0 at $t = 3$. Enumerating: $0 \to 1 \to 0 \to 0$, $0 \to 0 \to 1 \to 0$, $0 \to 1 \to 1 \to 0$, and $0 \to 0 \to 0 \to 0$. The associated probabilities are $pq(1-p)$, $(1-p)pq$, $p(1-p-q)q$ and $(1-p)^3$. The probability Willie is in the road is then $1 - 2pq(1-p) - pq(1-p-q) - (1-p)^3$.

   Or we simply multiply the the probability vector $[1, 0, 0, 0, 0]^\top$ by the transition matrix three times. This yields (laboriously)

$$\begin{bmatrix} \left((1-p)^2 + qp\right)(1-p) + ((1-p)\,q + q\,(1-p-q))\,p \\ (p\,(1-p) + (1-p-q)\,p)\,(1-p) + \left(2\,qp + (1-p-q)^2\right)p \\ p^2\,(1-p) + 2\,(1-p-q)\,p^2 \\ p^2 \\ 0 \end{bmatrix}$$

   and taking the middle three terms we have

$$\left[\ (p\,(1-p) + (1-q)\,p)\,(1-p) + (qp + (1-q)\,(1-p-q) + (2-p)\,p)\,p\ \right]$$

   I've not checked this answer against the first one, so beware algebra errors.

(c) *(10 points)* If the road is deserted and Willie's tape is infinitely long, please derive the probability distribution on Willie's position as the number of beats goes to infinity.

This is a typical birth death process, so we readily see that in the steady state we must have

$$\pi_i = \frac{p}{q}\pi_{i-1}$$

for $i = 1, 2, 3, 4$. *Using the constraint that the probabilities must sum to* 1 *we have*

$$\pi_i = \frac{1-\frac{p}{q}}{1-(\frac{p}{q})^5}\left(\frac{p}{q}\right)^i$$

(d) *(10 points)* Now suppose that at each beat a car has probability $r > 0$ of coming down the road and crushing Willie. Assume that this event is independent from beat to beat (iid). Also assume that Willie decides to move just after each beat but a car comes exactly on the beat – thus, Willie can never be crushed when he starts from a curb position. Sketch the Markov chain associated with this new system assuming that if Willie is crushed, he never moves again. Is this Markov chain ergodic? Why/why not? Does it have a unique steady state distribution?

We can either enumerate the "crush states" by having separate states associated with being crushed from positions 1,2 and 3, or we can have a single crush state. In all cases, Willie is eventually crushed, so the only nonzero probabability states as time marches toward infinity are the crush states.

In the latter (single crush state) there will be a unique steady state since eventually Willie gets crushed.

In neither case is the chain ergodic since not all states are recurrent (note that none are periodic so a-periodicity does not matter in this case).

The transtition equations for the finest grain case are

$$\pi(n{+}1) = \begin{bmatrix} 1-p & q(1-r) & 0 & 0 & 0 & 0 & 0 & 0 \\ p & (1-r)(1-q-p) & (1-r)q & 0 & 0 & 0 & 0 & 0 \\ 0 & (1-r)p & (1-r)(1-q-p) & (1-r)q & 0 & 0 & 0 & 0 \\ 0 & 0 & (1-r)p & (1-r)(1-q-p) & q & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & r & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & r & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & r & 0 & 0 & 0 & 1 \end{bmatrix} \pi(n)$$

where the final three states are the crush states associated with positions 1, 2 and 3 respectively. The transition equations for the combined crush state is

$$\pi(n{+}1) = \begin{bmatrix} 1-p & q(1-r) & 0 & 0 & 0 & 0 \\ p & (1-r)(1-q-p) & (1-r)q & 0 & 0 & 0 \\ 0 & (1-r)p & (1-r)(1-q-p) & (1-r)q & 0 & 0 \\ 0 & 0 & (1-r)p & (1-r)(1-q-p) & q & 0 \\ 0 & 0 & 0 & (1-r)p & 1-q & 0 \\ 0 & r & r & r & 0 & 1 \end{bmatrix} \pi(n)$$

In the former case, there is obviously more than one stationary probability distribution. Consider that there are three different stationary distributions associated with starting in each crushed state and therefore an infinite number of potential stationary distributions associated with starting distributions of the crush states. Slightly more sophisticated arguments based on asymmetry and the higher probability of being trapped in certain trap states depending on the initial position could also be used.

(e) *(10 points)* Derive an expression (and simplify it if possible) for the probability that Willie is still uncrushed after n beats assuming he started in position 0. What can you say quantitatively about the probability that Willie remains uncrushed as $n \to \infty$

Let $\mathbf{A}$ *be the state transition matrix appropriate for each of the above cases. The probability Willie is uncrushed after* n *beats is*

$$1 - [0, 0, 0, 0, 0, 1]\mathbf{A}^n \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

for the single crush state and

$$1 - [0, 0, 0, 0, 0, 1, 1, 1]\mathbf{A}^n \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

for the finest grain case.

2. *(50 points)* **Rutgera and Her True Love:** Rutgera Univera, the world famous Rutgers University graduate student has fallen in love with Jake her lab partner. Unfortunately, Jake does not return her ardor and she has engaged the services of an archer to shoot a love arrow into his heart. The archer operates under cover of night and has a probability p of shooting the arrow properly into Jake's heart.

   If Jake is NOT in love with Rutgera, then during the lab the number of words w he speaks to her follows a geometric distribution $p_W(w) = (1 - r)r^w$. If he is in love with Rutgera he speaks with distribution $p_W(w) = (1 - r/2)(r/2)^w$ (he's not a shy sort when he's in love).

   At the end of the next day's lab Rutgera will either declare her own love with a big fat kiss – or not. If Jake is not in love with her and she kisses him, it will cost her 10 happiness units in embarrassment. If he is in love with her and she kisses him, she will gain 10 happiness units. There is no cost associated with NOT kissing Jake.

Please develop an optimum decision rule based on the number of words spoken by Jake which Rutgera can use to maximize her happiness. Also derive an expression for Rutgera's expected value of happiness after the lab using the optimum decision rule.

This is a hypothesis testing problem. The null hypothesis is Jake was not shot and is therefore not in love. This happens with probability $1-p$. *The word distribution associated with this hypothesis is* $p_W(w) = (1-r)(r)^w$. *Likewise the positive hypothesis that Jake is in love happens with probability* p *and the associated word distribution is* $p_W(w) = (1-r/2)(r/2)^w$.

If $\mathcal{W}$ *is the set of observed* W *for which Rutgera kisses Jake then the cost metric is*

$$Q = 10\left[\sum_{w\in\mathcal{W}} p(1-r/2)(r/2)^w - (1-p)(1-r)(r)^w\right]$$

Clearly we want to choose $\mathcal{W}$ *so that* $p(1-r/2)(r/2)^w - (1-p)(1-r)(r)^w > 0$ *Rearranging we have*

$$-w \quad \underset{\textit{NO KISS}}{\overset{\textit{KISS}}{\gtrless}} \quad \log_2\left[\frac{(1-p)(1-r)}{p(1-r/2)}\right]$$

As for the expected value of Rutgera's happiness, you already derived it – it's the cost metric.

$$Q = 10\left[\sum_{w=0}^{\lfloor -\log_2\left[\frac{(1-p)(1-r)}{p(1-r/2)}\right]\rfloor} p(1-r/2)(r/2)^w - (1-p)(1-r)(r)^w\right]$$

3. *(30 points)* **Fun With Correlation:** A certain communication system has two observables of the form

$$r_1 = b + w_1$$
$$r_2 = b + w_2$$

where w_1 and w_1 are zero mean jointly guassian random variables with variance 1. The correlation coefficient between w_1 and w_2 is $\rho = -1$. b is a random variable which takes on values 1 and -1 equiprobably.

(a) *(10 points)* Please derive a receiver which uses r_1 and r_2 and makes a decision about the value of b with minimum probability of error. Calculate the probability of error of your receiver.

If two jointly Gaussian zero mean random variables with the same variance have correlation -1, *then* $w_1 = -w_2$. *If you tried to show this by simply using the expression for joint distributions of jointly Gaussian random variables, you ran into singularity problems. You could have used limiting arguments (find the distribution of* w_1 *conditioned on knowing* w_2*) but that's more work than necessary (though if you did it that way and got it right, no problem).*

The easiest way is to notice that when $\rho = -1$, $E((w_1+w_2)^2) = 0$ *which implies since* $E(w_1 + w_2) = 0$ *that the variance of* $w_1 + w_2$ *is zero. This only happens if* $w_1 + w_2$ *is deterministic which then implies that* $w_1 + w_2 = 0$ *or* $w_1 = -w_2$. *Notice that the jointly Gaussian assumption is not necessary if you take this route – and this is a result worth remembering.*

In any case, the optimum receiver simply adds r_1 *and* r_2 *to obtain* $2b$. *Detection is then error free since the noise terms* w_1 *and* w_2 *cancel.*

(b) *(20 points)* How does your answer change if $\rho = 0$? How about $\rho = 1$

If $\rho = 1$, then $w_1 = w_2$ and r_1 and r_2 are identical observations. To show this you use the argument in part (a) with $w_1 - w_2$. So we can essentially throw out either r_1 or r_2 this becomes the usual binary detection problem. Since $b = \pm 1$ equiprobably, we know that the optimal detector is

$$r_1 \underset{say\ b=-1}{\overset{say\ b=1}{\gtrless}} 0$$

and probability of error is

$$P_e = \int_0^\infty \sqrt{\frac{1}{2\pi}} e^{-x^2/2} dx$$

If $\rho = 0$ then we have two independent observationsw with independent noise. That means that

$$f_{R_1,R_2|b=1}(r_1, r_2) = \frac{1}{2\pi} e^{-(r_1-1)^2/2-(r_2-1)^2/2}$$

and

$$f_{R_1,R_2|b=1}(r_1, r_2) = \frac{1}{2\pi} e^{-(r_1+1)^2/2-(r_2+1)^2/2}$$

The cost metric is probability of error so we end up with the LRT

$$e^{-(r_1-1)^2/2-(r_2-1)^2/2+(r_1+1)^2/2-(r_2+1)^2/2} \underset{say\ b=-1}{\overset{say\ b=1}{\gtrless}} 1$$

which reduces to

$$r_1 + r_2 \underset{say\ b=-1}{\overset{say\ b=1}{\gtrless}} 0$$

The decision statistic is then $r = r_1 + r_2 = 2b + w_1 + w_2 = 2b + w$ where w is a zero mean Gaussian random variable with variance 2. We ran rewrite as $s = x + z$ where $x = 2b/2 = \pm 1$ and $z = (w_1 + w_2)/2$ and therefore has variance $1/2$. Thus, the probability of error is

$$P_e = \int_0^\infty \sqrt{\frac{1}{2\pi}} e^{-x^2} dx$$

4. *(20 points)* **Orthogonal Projections:** Suppose we are given a zero mean stationary white noise process $n(t)$ and derive two random variables

$$N_c = \int_0^1 n(t) \cos 2\pi t dt$$

and

$$N_s = \int_0^1 n(t) \sin 2\pi t dt$$

(a) *(10 points)* What are the means and variances of N_c and N_s? What is the joint distribution on N_c and N_s? *Means are zero, variances are $N_0/2$ if the spectral intensity of the noise process is N_0. We know that N_s and N_c are jointly Gaussian. The correlation is zero (the sine and cosine cancel out with integration over a single period), so they are independent Gaussian random variables with zero mean and variances $N_0/2$. The joint distribution is the product of the marginals.*

(b) *(10 points)* For functions $\{x_i(t)\}$, $i = 1, 2, ...$ which have

$$\int_a^b x_i(t)x_j(t)dt = \delta_{ij}$$

on some interval $[a, b]$, find a relationship between the random variables

$$N_i = \int_a^b n(t)x_i(t)dt$$

when $n(t)$ is a zero mean stationary white Gaussian process.

The N_i are jointly Guassian. The means are again zero. The covariances are zero for $i \neq j$ and N_0 for $i = j$. Don't believe?

$$E[N_i N_j] = \int_a^b \int_a^b E[n(t)n(\tau)]x_i(t)x_j(\tau)dtd\tau = N_0 \int_a^b x_i(t)x_j(t)dt = N_0\delta_{ij}$$

The random variables are therefore IID Gaussian.

5. *(50 points)* **An Interesting Sampling Theorem:** We all know that for a waveform $x(t)$ bandlimited to $\pm W$ it suffices to sample $x(t)$ with period $T = 1/2W$ in order to reconstruct $x(t)$ as

$$x(t) = \sum_n \frac{\sin 2\pi W(t - nT)}{2\pi W(t - nT)} x(nT)$$

Now suppose that $x(t)$ is sampled according to a Poisson process with rate λ where the t_i are the arrival times. Specifically we have samples

$$x_i = x(t_i) = x(\sum_{j=-\infty}^{i} \tau_i)$$

where $\tau_i = t_i - t_{i-1}$ and

$$f_{\tau_i}(\tau_i) = \lambda e^{-\lambda\tau_i}$$

Show that so long as $\lambda > 2W$, we may always reconstruct the signal $x(t)$ with zero mean square error with suitable $\phi_i(t)$ as

$$\lim_{A\to\infty} \frac{1}{2A} \int_{-A}^{A} \left(x(t) - \sum_i x_i\phi_i(t - t_i) \right)^2 dt = 0$$

This problem should have been spoon-fed to you with steps parceled out to guide you. So it will not count toward your grade. If you did something even halfway reasonable, it'll bump you up (after the grade boundaries are established).

The solution is a bit laborious but here's an outline.

We can represent $x(t)$ as

$$x(t) = \sum_n \frac{\sin 2\pi W(t - nT)}{2\pi W(t - nT)} x(nT)$$

and therefore from these equally-spaced samples $x(nT)$ we can obtain any values of $x(t)$ we desire. Another way to put this is that the $\{x(nT)\}$ COMPLETELY REPRESENT $x(t)$. So one way to approach the problem is to figure out how many irregular samples we need per unit time to be sure we can reconstruct the set $\{x(nT)\}$.

To do this, we proceed in two steps. First we start with bandlimited periodic functions. We show that for period Q, any set of Q/T different samples on a period will allow us to obtain the $\{x(nT)\}$. We then extend the period towards infinity and show by the strong law of large numbers that for $\lambda > 1/T$, the number of samples in a period is greater than Q/T with probability 1, thus completing the proof.

There are probably other ways, but I can't think of any off the top of my head. Only three people attempted the problem and only one person said something reasonable – and that was instinctive and heuristic (about the average number of samples). Regardless, it's a neat theorem and good to remember (theoretically). In terms of practical use, I think it's mostly useless – especially once you add a spot of noise.