18.177: Problem Set #3 (solutions)

1. The f-process has entropy strictly less than $\log 2$, while the g-process has entropy exactly $\log 2$. Let $y_i = f(T^i x)$, $z_i = g(T^i x)$. If $y_{-2} = \mathbf{1}$ and $y_{-1} = \mathbf{0}$ (which occurs with probability $\frac{1}{8}$), y_0 must be $\mathbf{0}$ (so that the conditional entropy of y_0 given $y_{-2} = \mathbf{1}$ and $y_{-1} = \mathbf{0}$ is 0), whereas in no case can the conditional entropy of y_0 given $\ldots, y_{-2}, y_{-1}$ be greater than $\log 2$ (since there are only two possible values for y_0). Hence the entropy of the y-process is at most $\frac{1}{8}(\log 1) + \frac{7}{8}(\log 2) = \frac{7}{8}\log 2 < \log 2$. (Alternatively, note that it is impossible to have $y_{n-1} = \mathbf{1}$, $y_n = \mathbf{0}$, and $y_{n+1} = \mathbf{1}$, so that the number of possible y-strings of length $3k$ is at most 7^k; it follows that $H((y)_0^{3k-1}) \leq \log 7^k$, so that, dividing by $3k$ and taking the limit, we get $h(T, f) \leq \frac{1}{3}\log 7 < \log 2$.) In contrast, the conditional distribution of z_0 given $\ldots, z_{-2}, z_{-1}$ is a two-point measure with probability distribution $(\frac{1}{2}, \frac{1}{2})$, so that $h(T, g) = H(z_0 | \ldots, z_{-2}, z_{-1}) = \log 2$.

2. (a) Yes, since $P \vee TP \vee T^2P = T^2(P \vee T^{-1}P \vee T^{-2}P)$. (b) No; let (X, T, Q) be a $(\frac{1}{2}, \frac{1}{2})$ i.i.d. process, with Q the natural time-0 partition, and let $P = Q \vee TQ \vee T^3Q$. Then $H(P \vee TP \vee T^3P) = \log 32$ while $H(P \vee T^{-1}P \vee T^{-3}P) = \log 64$. (c) As in part (a), we have $H(P \vee TP \vee \ldots \vee T^{n-1}P) = H(P \vee T^{-1}P \vee \ldots \vee T^{-(n-1)}P)$, so that, dividing by n and taking the limit, we have $h(T, P) = h(T^{-1}, P)$.

3. If P is any partition of the space X, then, defining $Q = P \vee TP$, we have $H(Q \vee T^2Q \vee \ldots \vee (T^2)^{n-1}Q) = H(P \vee TP \vee \ldots \vee T^{2n-1}P)$, so that (dividing by n taking the limit) we get $h(T^2, Q) = 2h(T, P)$. Since P is arbitrary, we have $h(T^2) = \sup_Q h(T^2, P) \geq \sup_P 2h(T, P) = 2h(T)$. On the other hand, $H(P \vee T^2P \vee \ldots \vee (T^2)^{n-1}P) \leq H(P \vee TP \vee \ldots \vee T^{2n-1}P)$, so that $h(T^2, P) \leq 2h(T, P) \leq 2h(T)$. Since P is arbitrary, $h(T^2) = \sup_P h(T^2, P) \leq 2h(T)$. Equality follows. (It is also possible to solve this problem more easily by noting that if P is a generator for T then $P \vee TP$ is a generator for T^2.)

4. Every T-invariant measure μ has a unique decomposition into ergodic measures, namely $\mu = \int_E m \, d\lambda(m)$, where E is the set of ergodic T-invariant measures on X and λ is some probability distribution on E. To prove that $h_\mu(T, Q)$ is affine in μ it is enough to prove that $h_\mu(T, Q) = \int_E h_m(T, Q) \, d\lambda(m)$. We can write the left and right hand sides of this equation as $H_\mu(Q | (Q)_{-\infty}^{-1})$

and $H_\mu(Q|(Q)_{-\infty}^{-1} \vee \mathcal{E})$, respectively, where $\mathcal{E}$ is the σ-algebra of invariant sets, which you may want to imagine (for intuitive purposes) as being associated with a random variable that takes its values in the set of ergodic components of μ, with distribution governed by λ. However, in the case where Q is a generator, the ergodic theorem tells us that $\mathcal{E}$ is a sub-σ-algebra of $(Q)_{-\infty}^{-1}$ modulo sets of measure 0 (because the statistics of $(Q)_{-\infty}^{-1}(x)$ almost surely discriminate which ergodic component x lies in), so that $H_\mu(Q|(Q)_{-\infty}^{-1}) = H_\mu(Q|(Q)_{-\infty}^{-1} \vee \mathcal{E})$; if Q is not a generator, then the same reasoning applies if one replaces $\mathcal{E}$ by the σ-algebra of $(Q)_{-\infty}^{\infty}$-measurable T-invariant sets. Hence $h_\mu(T,Q) = \int_E h_m(T,Q)\ d\lambda(m)$, so that $h_\mu(T,Q)$ is affine in μ, as claimed. Taking Q to be the universal generator P, we find that $h_\mu(T)$ is affine in μ as well.

(Alternative argument, given by David Gamarnik: It suffices to prove that $h_\mu(P,T) = t h_{\mu_1}(P,T) + (1-t) h_{\mu_2}(P,T)$ whenever $\mu = t\mu_1 + (1-t)\mu_2$ with $0 \leq t \leq 1$. Make two copies X_1, X_2 of the underlying space and model μ as a measure on the space $\overline{X} = X_1 \cup X_2$ with $\mu(X_1) = t$, $\mu(X_2) = 1-t$, $\mu|_{X_1} = \mu_1$, and $\mu|_{X_2} = \mu_2$. $\overline{X}$ comes with a partition $I = \{X_1, X_2\}$ and a partition that lifts P to $\overline{X}$ (call it $\overline{P}$). We have $H_\mu(\overline{P} \vee \ldots \vee T^{-n}\overline{P}) = t H_{\mu_1}(\overline{P} \vee \ldots \vee T^{-n}\overline{P}) + (1-t) H_{\mu_1}(\overline{P} \vee \ldots \vee T^{-n}\overline{P}) + (H_\mu(I) - H_\mu(I|\overline{P} \vee \ldots \vee T^{-n}\overline{P}))$. The third term on the right hand side lies between 0 and $H_\mu(I)$, and in particular, it is bounded; hence, when we divide by n and take the limit, it goes away, yielding the desired equality for $\overline{P}$.)

We have already seen in class that these quantities are *not* continuous in μ. For instance, let μ_n be a measure that assigns equal measure to all 2^n infinite binary sequences that are periodic with period n. Then $h_{\mu_n}(T,P) = 0$ for all n, even though the μ_n's converge to the $(\frac{1}{2}, \frac{1}{2})$ i.i.d. measure, for which $h(T,P) = \log 2$.

5. Yes. As was discussed in class, we have $h(T,P) \leq h(T,P') + H(P|P')$, which (applying symmetry) yields $|h(T,P) - h(T,P')| \leq H(P'|P) + H(P|P')$. On the other hand, $\mu(P \triangle P_n) \to 0$ implies $H(P'|P) + H(P|P') \to 0$.

6. No. For any $\epsilon > 0$, we can make $h(T,P)$ change by as much as we like merely by re-defining the map T on the top level $T^{n-1}A$ and left-over set $X \setminus \bigcup_{i=0}^{n-1} T^i A$ of a Rokhlin tower, such that these two sets taken together have total measure less than ϵ. (Cf. problem #5 on the second problem set.)

7. (sketch) Given a probability distribution on $\mathbf{Z}/n$, one can change the prob-

abilities of the outcomes $\frac{i}{n}$, $\frac{i+1}{n}$, $\frac{i+2}{n}$, and $\frac{i+3}{n}$ by t, $-3t$, $3t$, and $-t$, respectively, without affecting the sum of the probabilities, the expectation of the random variable, or the variance of the random variable. Such an alteration changes the entropy of the random variable (defined as $\sum_i p(\frac{i}{n}) \log 1/p(\frac{i}{n})$) by $\log p(\frac{i}{n}) - 3 \log p(\frac{i+1}{n}) + 3 \log p(\frac{i+2}{n}) - \log p(\frac{i+3}{n})$. If the entropy is to be at a maximum, this expression must vanish for all i. This implies that $p(\frac{i-1}{n})p(\frac{i+1}{n})/p(\frac{i}{n})^2$ must be constant (independent of i). (One can also use Lagrange multipliers, though in this infinite-dimensional setting some extra justification is required. In any case, a fully fleshed-out proof requires a lemma that rules out the situation in which some of the outcomes have vanishing probability.) It follows from this that $p(\frac{i}{n})$ is given by a function of the form Ae^{Bi^2} for suitable constants A, B. B can be found explicitly in terms of n; A cannot, but it is merely a normalizing factor that ensures that the probabilities sum to 1. Taking the limit as n goes to infinity yields the Gaussian distribution with mean 0 and variance 1.