2004 Spring Midterm for Information Theory

1. (10 pt) A uniquely decodable variable-length code with binary code alphabet $\{0,1\}$ and with 6 codewords having lengths $\ell_0, \ell_1, \ell_2, \ell_3, \ell_4, \ell_5$. Give an example of Huffman coding that equates the Kraft inequality, i.e., $\sum_{m=0}^{6} 2^{-\ell_m} = 1$.
   (Note: You need to specify the probabilities of each of the 6 outcomes, which results in a Huffman code that equates the Kraft inequality.)
   (Hint: Binary tree.)

   Answers: Code $\{0, 10, 110, 1110, 11110, 11111\}$ for probabilities, 1/2, 1/4, 1/8, 1/16, 1/32, 1/32. □

2. (a) (10 pt) For channel input X^n and channel output Y^n with joint distribution P_{X^n,Y^n}, write down the joint δ-typical set for this discrete memoryless channel.
   (Hint: Suppose P_{X^n,Y^n} is the joint distribution that achieves the channel capacity.)

   (b) (10 pt) For a given channel code $\mathcal{C} \triangleq \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_M\}$, specify the typical-set-based decoder used in the proof of Shannon's channel coding forward theorem in text.

   Answers: See slides I:4-14 and I:4-15. □

3. (10 pt)(Fano's inequality) Let X and Y be two random variables, correlated in general, with values in $\mathcal{X}$ and $\mathcal{Y}$, respectively, where $\mathcal{X}$ is finite but $\mathcal{Y}$ can be an infinite set. Let $\hat{x} \triangleq g(y)$ be an estimate of x from observing y. Define the probability of estimating error as
   $$P_e \triangleq \Pr\{g(Y) \neq X\}.$$
   Then for any estimating function $g(\cdot)$,
   $$H_b(P_e) + P_e \cdot \log(|\mathcal{X}| - 1) \geq H(X|Y),$$
   where $H_b(t) \triangleq -t \cdot \log t - (1-t) \cdot \log(1-t)$ is the binary entropy function.

   Answer: See slide I:4-28. □

4. (10 pt) Prove that the channel capacity of the binary-input-binary-output Z-channel is equal to:
   $$C = H_b\left(\frac{1}{1+e^{H_b(\epsilon)/\epsilon}}\right) - \frac{1}{\epsilon(1+e^{H_b(\epsilon)/\epsilon})} \cdot H_b(\epsilon)\text{nats/channel usage}$$
   where $\Pr\{Y=0|X=0\} = 1$ and $\Pr\{Y=1|X=1\} = \epsilon \in (0,1)$.
   (Hint: Let $p = \Pr\{X=1\}$. Compute $I(X;Y)$ in terms of $H(Y) - H(Y|X)$. Use $\partial H_b(x)/\partial x = \log((1-x)/x)$ to obtain the input distribution that maximizes the mutual information, where $H_b(x) = -x\log(x) - (1-x)\log(1-x)$.)

Answers: $H(Y|X) = (1-p)H(Y|X=0) + pH(Y|X=1) = p \cdot H_b(\epsilon)$ nats. So, $I(X;Y) = H(Y) - H(Y|X) = H(Y) - p \cdot H_b(\epsilon) = H_b(p\epsilon) - p \cdot H_b(\epsilon)$ nats. This gives that

$$\frac{\partial I(X;Y)}{\partial p} = \epsilon \log \frac{1-p\epsilon}{p\epsilon} - H_b(\epsilon).$$

By letting $\partial I(X;Y)/\partial p = 0$, we obtain

$$p^* = \frac{1}{\epsilon(1+e^{H_b(\epsilon)/\epsilon})}.$$

Therefore,

$$C = H_b\left(\frac{1}{1+e^{H_b(\epsilon)/\epsilon}}\right) - \frac{1}{\epsilon(1+e^{H_b(\epsilon)/\epsilon})} \cdot H_b(\epsilon)\text{nats/channel usage.}$$

□

5. (a) (10 pt) Derive the differential entropy of a source Y with $c \cdot e^{-c|y|}/2$ for $y \in \Re$.
   (b) (10 pt) Prove that the source Y with pdf $c \cdot e^{-c|y|}/2$ for $y \in \Re$ has the largest differential entropy among all sources with the same support (i.e., $\Re$) and the same first absolute moment (i.e., $E[|Y|]$).

Answers:

(a)

$$\begin{aligned}\int_{\Re} \frac{c}{2} e^{-c|y|} \left[-\log \frac{c}{2} + c|y|\right] dy &= -\log \frac{c}{2} + cE[|Y|] \\ &= 1 + \log(2) - \log(c).\end{aligned}$$

(b) Let $p(\cdot)$ be the pdf of a continuous source X with support $\Re$ and $E[|X|] = E[|Y|]$. Let $\phi(y) = ce^{-c|y|}/2$. Observe that

$$\begin{aligned}-\int_{\Re} \phi(y) \log \phi(y) dy &= \int_{\Re} \phi(y) \left[-\log \frac{c}{2} + c|y|\right] dy \\ &= -\log \frac{c}{2} + cE[|Y|] \\ &= -\log \frac{c}{2} + cE[|X|] \\ &= \int_{\Re} p(x) \left[-\log \frac{c}{2} + c|x|\right] dx \\ &= -\int_{\Re} p(x) \log \phi(x) dx.\end{aligned}$$

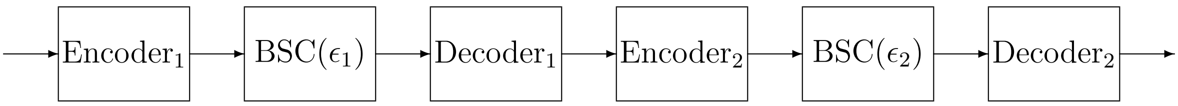
Hence,

$$\begin{aligned}
h(Y) - h(X) &= -\int_{\Re} \phi(y) \log \phi(y) dy + \int_{\Re} p(x) \log p(x) dx \\
&= -\int_{\Re} p(x) \log \phi(x) dx + \int_{\Re} p(x) \log p(x) dx \\
&= \int_{\Re} p(x) \log \frac{p(x)}{\phi(x)} dx \\
&\geq \int_{\Re} p(x) \left(1 - \frac{\phi(x)}{p(x)}\right) dx \quad \text{(fundamental inequality)} \\
&= \int_{\Re} (p(x) - \phi(x))\, dx \\
&= 0,
\end{aligned}$$

with equality holds if, and only if, $p(x) = \phi(x)$ for all $x \in \Re$. □

6. (20 pt) Assume that the two memoryless BSC channels below are independent. Determine the maximum reliable transmission rates respectively for the following two systems, in which System B allows an intermediate re-coding to help improving the system performance.
(Hint: The channel capacity of BSC(ϵ) is given by $1 - H_b(\epsilon)$ bits/channel usage, where $H_b(\epsilon) = -\epsilon \log_2 \epsilon - (1 - \epsilon) \log_2 (1 - \epsilon)$ is the binary entropy function.)

System A



System B

Answers: For System A, the concatenation of BSC(ϵ_1) and BSC(ϵ_2) becomes BSC($(1-\epsilon_1)\epsilon_2 + (1-\epsilon_2)\epsilon_1$). So, by Shannon's channel coding theorem, the maximum reliable transmission rate is $1 - H_b((1-\epsilon_1)\epsilon_2 + (1-\epsilon_2)\epsilon_1)$ bits/channel usage.

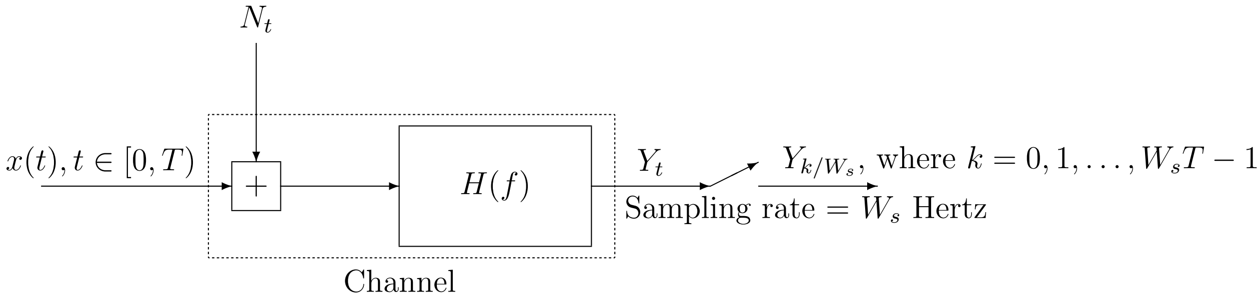
For System B, $P_{\text{error,SystemB}} = 1-(1-P_{\text{error,Decoder1}})(1-P_{\text{error,Decoder2}})$. Hence, $P_{\text{error,SystemB}}$ can be made arbitrarily small, if both $P_{\text{error,Decoder1}}$ and $P_{\text{error,Decoder2}}$ can be made arbitrarily small. Also, $P_{\text{error,SystemB}}$ is bounded away from zero, if $P_{\text{error,Decoder1}}$ or $P_{\text{error,Decoder2}}$ is bounded away from zero. Consequently, the maximum reliable transmission rate is given by $\min\{1 - H_b(\epsilon_1), 1 - H_b(\epsilon_2)\}$ bits/channel usage.

□

7. (10 pt) Below is a bandlimited waveform channel, where

$$H(f) = \begin{cases} \dfrac{1}{\sqrt{2W}}, & \text{for } -W \text{ (Hertz)} < f < W \text{ (Hertz)}; \\ 0, & \text{otherwise.} \end{cases}$$

Assume that W_sT is an integer.



Now suppose $X_t = 0$ for every t and N_t is a white noise, find the condition (i.e., the relation between W and W_s) under which Y_{i/W_s} and Y_{j/W_s} are uncorrelated for every $i \neq j$.

Answers:

In this case, $Y_t = N_t * H(t)$.

$$
\begin{aligned}
E[Y_{i/W_s} Y_{j/W_s}] &= E\left[\left(\int_{\Re} h(\tau) N_{(i/W_s)-\tau} d\tau\right)\left(\int_{\Re} h(\tau') N_{(j/W_s)-\tau'} d\tau'\right)\right] \\
&= \int_{\Re}\int_{\Re} h(\tau)h(\tau') E\left[N_{(i/W_s)-\tau} N_{(j/W_s)-\tau'}\right] d\tau' d\tau \\
&= \int_{\Re}\int_{\Re} h(\tau)h(\tau') \frac{N_0}{2}\delta\left(\frac{i}{W_s} - \frac{j}{W_s} - \tau + \tau'\right) d\tau' d\tau \\
&= \frac{N_0}{2}\int_{\Re} h(\tau) h(\tau - (i-j)/W_s) d\tau \\
&= \frac{N_0}{2}\int_{\Re}\left(\int_{-W}^{W} \frac{1}{\sqrt{2W}} e^{j2\pi f\tau} df\right)\left(\int_{-W}^{W} \frac{1}{\sqrt{2W}} e^{j2\pi f'(\tau-(i-j)/W_s)} df'\right) d\tau \\
&= \frac{N_0}{4W}\int_{-W}^{W}\int_{-W}^{W}\left(\int_{\Re} e^{j2\pi(f+f')\tau} d\tau\right) e^{-j2\pi f'(i-j)/W_s} df' df \\
&= \frac{N_0}{4W}\int_{-W}^{W}\int_{-W}^{W} \delta(f+f') e^{-j2\pi f'(i-j)/W_s} df' df \\
&= \frac{N_0}{4W}\int_{-W}^{W} e^{j2\pi f(i-j)/W_s} df \\
&= \frac{N_0}{2}\frac{\sin(2\pi W(i-j)/W_s)}{2\pi W(i-j)/W_s} \\
&= \begin{cases} N_0/2, & \text{if } i = j; \\ \dfrac{N_0}{2}\dfrac{\sin(2\pi W(i-j)/W_s)}{2\pi W(i-j)/W_s}, & \text{if } i \neq j. \end{cases}
\end{aligned}
$$

Hence, it requires that for any $i \neq j$, $(W/W_s)(i-j) = k/2$ for some $k = \pm 1, \pm 2, \pm 3, \ldots$, which is implied by "$W/W_s = k/2$ for some $k = 1, 2, 3, \ldots$" (or equivalently, $W_s = 2W/k$ for some $k = 1, 2, 3, \ldots$).