36) **Channel with memory.**
Consider the discrete memoryless channel $Y_i = Z_i X_i$ with input alphabet $X_i \in \{-1, 1\}$.

a) What is the capacity of this channel when $\{Z_i\}$ is i.i.d. with

$$Z_i = \begin{cases} 1, & p = 0.5 \\ -1, & p = 0.5 \end{cases} ? \tag{579}$$

Now consider the channel with memory. Before transmission begins, $Z$ is randomly chosen and fixed for all time. Thus $Y_i = Z X_i$.

b) What is the capacity if

$$Z = \begin{cases} 1, & p = 0.5 \\ -1, & p = 0.5 \end{cases} ? \tag{580}$$

36) **Channel with memory solution.**

a) This is a BSC with cross over probability 0.5, so $C = 1 - H(p) = 0$.

b) Consider the coding scheme of sending $X^n = (1, b_1, b_2, \dots, b_{n-1})$ where the first symbol is always a zero and the rest of the $n-1$ symbols are $\pm 1$ bits. For the first symbol $Y_1 = Z$, so the receiver knows $Z$ exactly. After that the receiver can recover the remaining bits error free. So in $n$ symbol transmissions $n$ bits are sent, for a rate $R = \frac{n-1}{n} \to 1$. The capacity $C$ is bounded by $\log |\mathcal{X}| = 1$, therefore the capacity is 1 bit per symbol.

1) **Differential entropy.** Evaluate the differential entropy $h(X) = -\int f \ln f$ for the following:

a) The exponential density, $f(x) = \lambda e^{-\lambda x}$ , $x \geq 0$.

b) The Laplace density, $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$.

c) The sum of $X_1$ and $X_2$, where $X_1$ and $X_2$ are independent normal random variables with means $\mu_i$ and variances $\sigma_i^2, i = 1, 2$.

1) *Differential Entropy.*

a) Exponential distribution.

$$\begin{aligned} h(f) &= -\int_0^\infty \lambda e^{-\lambda x}[\ln \lambda - \lambda x] dx & (690) \\ &= -\ln \lambda + 1 \text{ nats.} & (691) \\ &= \log \frac{e}{\lambda} \text{ bits.} & (692) \end{aligned}$$

b) Laplace density.

$$\begin{aligned} h(f) &= -\int_{-\infty}^\infty \frac{1}{2}\lambda e^{-\lambda|x|}[\ln \frac{1}{2} + \ln \lambda - \lambda|x|]\, dx & (693) \\ &= -\ln \frac{1}{2} - \ln \lambda + 1 & (694) \\ &= \ln \frac{2e}{\lambda} \text{ nats.} & (695) \\ &= \log \frac{2e}{\lambda} \text{ bits.} & (696) \end{aligned}$$

c) Sum of two normal distributions.
The sum of two normal random variables is also normal, so applying the result derived the class for the normal distribution, since $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

$$h(f) = \frac{1}{2} \log 2\pi e(\sigma_1^2 + \sigma_2^2) \text{ bits.} \tag{697}$$

2) **Concavity of determinants.** Let $K_1$ and $K_2$ be two symmetric nonnegative definite $n \times n$ matrices. Prove the result of Ky Fan [**?**]:

$$\mid \lambda K_1 + \overline{\lambda} K_2 \mid \geq \mid K_1 \mid^{\lambda} \mid K_2 \mid^{\overline{\lambda}}, \qquad \text{for } 0 \leq \lambda \leq 1, \;\; \overline{\lambda} = 1 - \lambda,$$

where $\mid K \mid$ denotes the determinant of $K$.
*Hint*: Let $\mathbf{Z} = \mathbf{X}_\theta$, where $\mathbf{X}_1 \sim N(0, K_1)$, $\mathbf{X}_2 \sim N(0, K_2)$ and $\theta = \text{Bernoulli}(\lambda)$. Then use $h(\mathbf{Z} \mid \theta) \leq h(\mathbf{Z})$.

2) *Concavity of Determinants.* Let $X_1$ and $X_2$ be normally distributed $n$-vectors, $\mathbf{X}_i \sim \phi_{K_i}(\mathbf{x})$, $i = 1, 2$. Let the random variable $\theta$ have distribution $\Pr\{\theta = 1\} = \lambda$, $\Pr\{\theta = 2\} = 1 - \lambda$, $0 \leq \lambda \leq 1$. Let $\theta$, $\mathbf{X}_1$, and $\mathbf{X}_2$ be independent and let $\mathbb{Z} = \mathbf{X}_\theta$. Then $\mathbb{Z}$ has covariance $K_Z = \lambda K_1 + (1 - \lambda) K_2$. However, $\mathbb{Z}$ will not be multivariate normal. However, since a normal distribution maximizes the entropy for a given variance, we have

$$\frac{1}{2} \ln(2\pi e)^n |\lambda K_1 + (1-\lambda) K_2| \geq h(\mathbb{Z}) \geq h(\mathbb{Z}|\theta) = \lambda \frac{1}{2} \ln(2\pi e)^n |K_1| + (1-\lambda) \frac{1}{2} \ln(2\pi e)^n |K_2| \;.$$

Thus

$$|\lambda K_1 + (1-\lambda) K_2| \geq |K_1|^{\lambda} |K_2|^{1-\lambda} \;, \tag{698}$$

as desired.

7) **Differential entropy bound on discrete entropy:** Let $X$ be a discrete random variable on the set $\mathcal{X} = \{a_1, a_2, \ldots\}$ with $\Pr(X = a_i) = p_i$. Show that

$$H(p_1, p_2, \ldots) \leq \frac{1}{2} \log(2\pi e) \left( \sum_{i=1}^{\infty} p_i i^2 - \left( \sum_{i=1}^{\infty} i p_i \right)^2 + \frac{1}{12} \right) . \tag{688}$$

Moreover, for every permutation $\sigma$,

$$H(p_1, p_2, \ldots) \leq \frac{1}{2} \log(2\pi e) \left( \sum_{i=1}^{\infty} p_{\sigma(i)} i^2 - \left( \sum_{i=1}^{\infty} i p_{\sigma(i)} \right)^2 + \frac{1}{12} \right) . \tag{689}$$

*Hint:* Construct a random variable $X'$ such that $\Pr(X' = i) = p_i$. Let $U$ be an uniform(0,1] random variable and let $Y = X' + U$, where $X'$ and $U$ are independent. Use the maximum entropy bound on $Y$ to obtain the bounds in the problem. This bound is due to Massey (unpublished) and Willems(unpublished).

7) *Differential entropy bound on discrete entropy*
Of all distributions with the same variance, the normal maximizes the entropy. So the entropy of the normal gives a good bound on the differential entropy in terms of the variance of the random variable.
Let $X$ be a discrete random variable on the set $\mathcal{X} = \{a_1, a_2, \ldots\}$ with

$$\Pr(X = a_i) = p_i. \tag{717}$$

$$H(p_1, p_2, \ldots) \leq \frac{1}{2}\log(2\pi e)\left(\sum_{i=1}^{\infty} p_i i^2 - \left(\sum_{i=1}^{\infty} i p_i\right)^2 + \frac{1}{12}\right). \tag{718}$$

Moreover, for every permutation $\sigma$,

$$H(p_1, p_2, \ldots) \leq \frac{1}{2}\log(2\pi e)\left(\sum_{i=1}^{\infty} p_{\sigma(i)} i^2 - \left(\sum_{i=1}^{\infty} i p_{\sigma(i)}\right)^2 + \frac{1}{12}\right). \tag{719}$$

Define two new random variables. The first, $X_0$, is an integer-valued discrete random variable with the distribution

$$\Pr(X_0 = i) = p_i. \tag{720}$$

Let $U$ be a random variable uniformly distributed on the range $[0, 1]$, independent of $X_0$. Define the continuous random variable $\tilde{X}$ by

$$\tilde{X} = X_0 + U. \tag{721}$$

It is clear that $H(X) = H(X_0)$, since discrete entropy depends only on the probabilities and not on the values of the outcomes. Now

$$\begin{aligned}
H(X_0) &= -\sum_{i=1}^{\infty} p_i \log p_i & (722)\\
&= -\sum_{i=1}^{\infty} \left(\int_i^{i+1} f_{\tilde{X}}(x)\, dx\right) \log\left(\int_i^{i+1} f_{\tilde{X}}(x)\, dx\right) & (723)\\
&= -\sum_{i=1}^{\infty} \int_i^{i+1} f_{\tilde{X}}(x) \log f_{\tilde{X}}(x)\, dx & (724)\\
&= -\int_1^{\infty} f_{\tilde{X}}(x) \log f_{\tilde{X}}(x)\, dx & (725)\\
&= h(\tilde{X}), & (726)
\end{aligned}$$

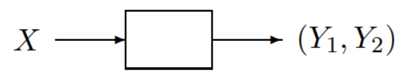
since $f_{\tilde{X}}(x) = p_i$ for $i \leq x < i + 1$.
Hence we have the following chain of inequalities:

$$\begin{aligned}
H(X) &= H(X_0) & (727)\\
&= h(\tilde{X}) & (728)\\
&\leq \frac{1}{2}\log(2\pi e)\mathrm{Var}(\tilde{X}) & (729)\\
&= \frac{1}{2}\log(2\pi e)\left(\mathrm{Var}(X_0) + \mathrm{Var}(U)\right) & (730)\\
&= \frac{1}{2}\log(2\pi e)\left(\sum_{i=1}^{\infty} p_i i^2 - \left(\sum_{i=1}^{\infty} i p_i\right)^2 + \frac{1}{12}\right). & (731)
\end{aligned}$$

Since entropy is invariant with respect to permutation of $p_1, p_2, \ldots$, we can also obtain a bound by a permutation of the $p_i$'s. We conjecture that a good bound on the variance will be achieved when the high probabilities are close together, i.e, by the assignment $\ldots, p_5, p_3, p_1, p_2, p_4, \ldots$ for $p_1 \geq p_2 \geq \cdots$.
How good is this bound? Let $X$ be a Bernoulli random variable with parameter $\frac{1}{2}$, which implies that $H(X) = 1$. The corresponding random variable $X_0$ has variance $\frac{1}{4}$, so the bound is

$$H(X) \leq \frac{1}{2}\log(2\pi e)\left(\frac{1}{4} + \frac{1}{12}\right) = 1.255 \text{ bits.} \tag{732}$$

2) **The two-look Gaussian channel.**




Consider the ordinary Gaussian channel with two correlated looks at X, i.e., $Y = (Y_1, Y_2)$, where

$$Y_1 = X + Z_1 \tag{780}$$

$$Y_2 = X + Z_2 \tag{781}$$

with a power constraint $P$ on $X$, and $(Z_1, Z_2) \sim \mathcal{N}_2(\mathbf{0}, K)$, where

$$K = \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}. \tag{782}$$

Find the capacity $C$ for

a) $\rho = 1$
b) $\rho = 0$
c) $\rho = -1$

2) *The two look Gaussian channel.*
It is clear that the input distribution that maximizes the capacity is $X \sim \mathcal{N}(0, P)$. Evaluating the mutual information for this distribution,

$$\begin{aligned} C_2 &= \max I(X; Y_1, Y_2) && (801)\\ &= h(Y_1, Y_2) - h(Y_1, Y_2|X) && (802)\\ &= h(Y_1, Y_2) - h(Z_1, Z_2|X) && (803)\\ &= h(Y_1, Y_2) - h(Z_1, Z_2) && (804) \end{aligned}$$

Now since

$$(Z_1, Z_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}\right), \tag{805}$$

we have

$$h(Z_1, Z_2) = \frac{1}{2}\log(2\pi e)^2|K_Z| = \frac{1}{2}\log(2\pi e)^2 N^2(1-\rho^2). \tag{806}$$

Since $Y_1 = X + Z_1$, and $Y_2 = X + Z_2$, we have

$$(Y_1, Y_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} P+N & P+\rho N \\ P+\rho N & P+N \end{bmatrix}\right), \tag{807}$$

and

$$h(Y_1, Y_2) = \frac{1}{2}\log(2\pi e)^2|K_Y| = \frac{1}{2}\log(2\pi e)^2(N^2(1-\rho^2) + 2PN(1-\rho)). \tag{808}$$

Hence the capacity is

$$\begin{aligned} C_2 &= h(Y_1, Y_2) - h(Z_1, Z_2) && (809)\\ &= \frac{1}{2}\log\left(1 + \frac{2P}{N(1+\rho)}\right). && (810) \end{aligned}$$

a) $\rho = 1$. In this case, $C = \frac{1}{2}\log(1 + \frac{P}{N})$, which is the capacity of a single look channel. This is not surprising, since in this case $Y_1 = Y_2$.

b) $\rho = 0$. In this case,

$$C = \frac{1}{2}\log\left(1 + \frac{2P}{N}\right), \tag{811}$$

which corresponds to using twice the power in a single look. The capacity is the same as the capacity of the channel $X \to (Y_1 + Y_2)$.

c) $\rho = -1$. In this case, $C = \infty$, which is not surprising since if we add $Y_1$ and $Y_2$, we can recover $X$ exactly.

Note that the capacity of the above channel in all cases is the same as the capacity of the channel $X \to Y_1 + Y_2$.