An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Aug. 29, 2016

Lecture #12A: Gaussian Channel

Outline of the lecture

- The Gaussian channel

Outline of the lecture

- The Gaussian channel
- Achievability of Gaussian channel capacity

Outline of the lecture

- The Gaussian channel
- Achievability of Gaussian channel capacity
- Converse to the coding theorem for Gaussian channels

Outline of the lecture

- The Gaussian channel
- Achievability of Gaussian channel capacity
- Converse to the coding theorem for Gaussian channels
- Bandlimited Gaussian channel

The Gaussian Channel

- The capacity of the Gaussian channel with power constraint P and noise variance N is given by

$$C = \max_{EX^2 \le P} I(X;Y) = \frac{1}{2} \log \left(1 + \frac{P}{N}\right)$$

where maximum is attained when $X \sim N(0, K)$.

The Gaussian Channel

- The capacity of the Gaussian channel with power constraint P and noise variance N is given by

$$C = \max_{EX^2 \le P} I(X;Y) = \frac{1}{2} \log \left(1 + \frac{P}{N}\right)$$

where maximum is attained when $X \sim N(0, K)$.

Proof:

- Capacity is given by

$$C = \max_{p(x): EX^2 \le P} I(X;Y)$$

The Gaussian Channel

- We can write I(X;Y) as

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z) \end{aligned}$$

The Gaussian Channel

- We can write I(X;Y) as

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z) \end{aligned}$$

- We have $h(Z) = \frac{1}{2}\log 2\pi e N$. And we also have

$$EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2 = P + N$$

The Gaussian Channel

- Thus we have

$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Z) \\
&\leq \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi e N \\
&= \frac{1}{2}\log\left(1+\frac{P}{N}\right)
\end{aligned}
$$

The Gaussian Channel

- Thus we have

$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Z) \\
&\leq \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi e N \\
&= \frac{1}{2}\log\left(1+\frac{P}{N}\right)
\end{aligned}
$$

- Maximum is attained when $X \sim N(0, P)$.

Achievability of Gaussian channel capacity

- The capacity of a Gaussian channel with power constraint P and noise variance N is

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right) \quad \text{bits per transmission}$$

Achievability of Gaussian channel capacity

- The capacity of a Gaussian channel with power constraint P and noise variance N is

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right) \quad \text{bits per transmission}$$

- **Proof:** *(Achievability).* We will use the same ideas as in the proof of the channel coding theorem in the case of discrete channels: namely, random codes and joint typicality decoding.

Achievability of Gaussian channel capacity

- The capacity of a Gaussian channel with power constraint P and noise variance N is

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \quad \text{bits per transmission}$$

- **Proof:** *(Achievability).* We will use the same ideas as in the proof of the channel coding theorem in the case of discrete channels: namely, random codes and joint typicality decoding.
- We will take into account the power constraint and the fact that the variables are continuous and not discrete.

Achievability of Gaussian channel capacity

- *Generation of the codebook* We wish to generate a codebook in which all the codewords satisfy the power constraint.

Achievability of Gaussian channel capacity

- *Generation of the codebook* We wish to generate a codebook in which all the codewords satisfy the power constraint.
- To ensure this, we generate the codewords with each element i.i.d. according to a normal distribution with variance $P - \epsilon$.

Achievability of Gaussian channel capacity

- *Generation of the codebook* We wish to generate a codebook in which all the codewords satisfy the power constraint.
- To ensure this, we generate the codewords with each element i.i.d. according to a normal distribution with variance $P - \epsilon$.
- Since for large n, $\frac{1}{n}\sum X_i^2 \to P - \epsilon$, the probability that a codeword does not satisfy the power constraint will be small.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Achievability of Gaussian channel capacity

- *Generation of the codebook* We wish to generate a codebook in which all the codewords satisfy the power constraint.
- To ensure this, we generate the codewords with each element i.i.d. according to a normal distribution with variance $P - \epsilon$.
- Since for large n, $\frac{1}{n}\sum X_i^2 \to P - \epsilon$, the probability that a codeword does not satisfy the power constraint will be small.
- Let $X_i(w), i = 1, 2, \ldots, n, w = 1, 2, \ldots, 2^{nR}$ be i.i.d. $\sim \mathcal{N}(0, P - \epsilon)$ forming codewords $X^n(1), X^n(2), \ldots, X^n(2^{nR}) \in \mathcal{R}^n$.

Achievability of Gaussian channel capacity

- *Encoding.* After the generation of the codebook, the codebook is revealed to both the sender and the receiver.

Achievability of Gaussian channel capacity

- *Encoding.* After the generation of the codebook, the codebook is revealed to both the sender and the receiver.
- To send the message index w, the transmitter sends the wth codeword $X^n(w)$ in the codebook.

Achievability of Gaussian channel capacity

- *Encoding.* After the generation of the codebook, the codebook is revealed to both the sender and the receiver.
- To send the message index w, the transmitter sends the wth codeword $X^n(w)$ in the codebook.
- *Decoding.* The receiver looks down the list of codewords $\{X^n(w)\}$ and searches for one that is jointly typical with the received vector.

Achievability of Gaussian channel capacity

- *Encoding.* After the generation of the codebook, the codebook is revealed to both the sender and the receiver.
- To send the message index w, the transmitter sends the wth codeword $X^n(w)$ in the codebook.
- *Decoding.* The receiver looks down the list of codewords $\{X^n(w)\}$ and searches for one that is jointly typical with the received vector.
- If there is one and only one such codeword $X^n(w)$, the receiver declares $\hat{W} = w$ to be the transmitted codeword. Otherwise, the receiver declares an error.

Achievability of Gaussian channel capacity

- *Encoding.* After the generation of the codebook, the codebook is revealed to both the sender and the receiver.
- To send the message index w, the transmitter sends the wth codeword $X^n(w)$ in the codebook.
- *Decoding.* The receiver looks down the list of codewords $\{X^n(w)\}$ and searches for one that is jointly typical with the received vector.
- If there is one and only one such codeword $X^n(w)$, the receiver declares $\hat{W} = w$ to be the transmitted codeword. Otherwise, the receiver declares an error.
- The receiver also declares an error if the chosen codeword does not satisfy the power constraint.

Achievability of Gaussian channel capacity

- *Probability of error.* Without loss of generality, assume that codeword 1 was sent. Thus, $Y^n = X^n(1) + Z^n$. We define the following events:

$$E_0 = \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j^2(1) > P \right\}$$

and

$$E_i = \{(X^n(i), Y^n) \text{ is in } A_\epsilon^{(n)}\}$$

Achievability of Gaussian channel capacity

- *Probability of error.* Without loss of generality, assume that codeword 1 was sent. Thus, $Y^n = X^n(1) + Z^n$. We define the following events:

$$E_0 = \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j^2(1) > P \right\}$$

and

$$E_i = \{(X^n(i), Y^n) \text{ is in } A_\epsilon^{(n)}\}$$

- Then an error occurs if E_0 occurs (the power constraint is violated) or E_1^c occurs (the transmitted codeword and the received sequence are not jointly typical) or $E_2 \cup E_3 \cup \cdots E_{2^{nR}}$ occurs (some wrong codeword is jointly typical with the received sequence).

Achievability of Gaussian channel capacity

- Let $\mathcal{E}$ denote the event $\hat{W} \neq W$ and let P denote the conditional probability given that $W = 1$. Hence,

$$Pr(\mathcal{E}|W=1) = P(\mathcal{E}) = P(E_0 \cup E_1^c \cup E_2 \cup E_3 \cup \cdots E_{2^{nR}})$$
$$\leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$

by the union of events bound for probabilities.

Achievability of Gaussian channel capacity

- Let $\mathcal{E}$ denote the event $\hat{W} \neq W$ and let P denote the conditional probability given that $W = 1$. Hence,

$$Pr(\mathcal{E}|W=1) = P(\mathcal{E}) = P(E_0 \cup E_1^c \cup E_2 \cup E_3 \cup \cdots E_{2^{nR}})$$
$$\leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$

by the union of events bound for probabilities.
- By the law of large numbers, $P(E_0) \rightarrow 0$ as $n \rightarrow \infty$.

Achievability of Gaussian channel capacity

- Let $\mathcal{E}$ denote the event $\hat{W} \neq W$ and let P denote the conditional probability given that $W = 1$. Hence,

$$Pr(\mathcal{E}|W=1) = P(\mathcal{E}) = P(E_0 \cup E_1^c \cup E_2 \cup E_3 \cup \cdots E_{2^{nR}})$$
$$\leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$

by the union of events bound for probabilities.

- By the law of large numbers, $P(E_0) \to 0$ as $n \to \infty$.
- Now, by the joint AEP (which can be proved using the same argument as that used in the discrete case), $P(E_1^c) \to 0$, and hence

$$P(E_1^c) \leq \epsilon \quad \text{for } n \text{ sufficiently large}$$

Achievability of Gaussian channel capacity

- Since by the code generation process, $X^n(1)$ and $X^n(i)$ are independent, so are Y^n and $X^n(i)$.

Achievability of Gaussian channel capacity

- Since by the code generation process, $X^n(1)$ and $X^n(i)$ are independent, so are Y^n and $X^n(i)$.
- Hence, the probability that $X^n(i)$ and Y^n will be jointly typical is $\leq 2^{-n(I(X;Y)-3\epsilon)}$ by the joint AEP.

Achievability of Gaussian channel capacity

- Since by the code generation process, $X^n(1)$ and $X^n(i)$ are independent, so are Y^n and $X^n(i)$.
- Hence, the probability that $X^n(i)$ and Y^n will be jointly typical is $\leq 2^{-n(I(X;Y)-3\epsilon)}$ by the joint AEP.
- Now let W be uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$ and consequently,

$$Pr(\mathcal{E}) = \frac{1}{2^{nR}} \sum \lambda_i = P_\epsilon^{(n)}$$

Achievability of Gaussian channel capacity

- Then

$$P_\epsilon^{(n)} = Pr(\mathcal{E}) = Pr(\mathcal{E}|W=1)$$
$$\leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$
$$\leq \epsilon + \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)}$$
$$= 2\epsilon + (2^{nR}-1)2^{-n(I(X;Y)-3\epsilon)}$$
$$\leq 2\epsilon + 2^{3n\epsilon}2^{-n(I(X;Y)-R)}$$
$$\leq 3\epsilon$$

for n sufficiently large and $R < I(X;Y) - 3\epsilon$.

Achievability of Gaussian channel capacity

- Then

$$P_\epsilon^{(n)} = Pr(\mathcal{E}) = Pr(\mathcal{E}|W=1)$$
$$\leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$
$$\leq \epsilon + \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)}$$
$$= 2\epsilon + (2^{nR}-1)2^{-n(I(X;Y)-3\epsilon)}$$
$$\leq 2\epsilon + 2^{3n\epsilon}2^{-n(I(X;Y)-R)}$$
$$\leq 3\epsilon$$

for n sufficiently large and $R < I(X;Y) - 3\epsilon$.

- This proves the existence of a good $(2^{nR}, n)$ code.

Achievability of Gaussian channel capacity

- Now choosing a good codebook and deleting the worst half of the codewords, we obtain a code with low maximal probability of error.

Achievability of Gaussian channel capacity

- Now choosing a good codebook and deleting the worst half of the codewords, we obtain a code with low maximal probability of error.
- In particular, the power constraint is satisfied by each of the remaining codewords (since the codewords that do not satisfy the power constraint have probability of error 1 and must belong to the worst half of the codewords).

Achievability of Gaussian channel capacity

- Now choosing a good codebook and deleting the worst half of the codewords, we obtain a code with low maximal probability of error.
- In particular, the power constraint is satisfied by each of the remaining codewords (since the codewords that do not satisfy the power constraint have probability of error 1 and must belong to the worst half of the codewords).
- Hence we have constructed a code that achieves a rate arbitrarily close to capacity.

Converse to the Coding Theorem for Gaussian Channels

- We proof that the capacity of a Gaussian channel is $C = \frac{1}{2}\log(1 + \frac{P}{N})$ by proving that rates $R > C$ are not achievable.

Converse to the Coding Theorem for Gaussian Channels

- We proof that the capacity of a Gaussian channel is $C = \frac{1}{2}\log(1 + \frac{P}{N})$ by proving that rates $R > C$ are not achievable.
- The proof is similar to the proof for the discrete channel. The main new ingredient is the power constraint.

Converse to the Coding Theorem for Gaussian Channels

- We proof that the capacity of a Gaussian channel is $C = \frac{1}{2}\log(1 + \frac{P}{N})$ by proving that rates $R > C$ are not achievable.
- The proof is similar to the proof for the discrete channel. The main new ingredient is the power constraint.
- **Proof:** We must show that if $P_e^{(n)} \to 0$ for a sequence of $(2^{nR}, n)$ codes for a Gaussian channel with power constraint P, then

$$R \leq C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$$

Converse to the Coding Theorem for Gaussian Channels

- Consider any $(2^{nR}, n)$ code that satisfies the power constraint, that is,

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \leq P$$

for $w = 1, 2, \ldots, 2^{nR}$.

Converse to the Coding Theorem for Gaussian Channels

- Consider any $(2^{nR}, n)$ code that satisfies the power constraint, that is,

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \leq P$$

for $w = 1, 2, \ldots, 2^{nR}$.
- Proceeding as in the converse for the discrete case, let W be distributed uniformly over $\{1, 2, \ldots, 2^{nR}\}$.

Converse to the Coding Theorem for Gaussian Channels

- Consider any $(2^{nR}, n)$ code that satisfies the power constraint, that is,
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \leq P$$
for $w = 1, 2, \ldots, 2^{nR}$.
- Proceeding as in the converse for the discrete case, let W be distributed uniformly over $\{1, 2, \ldots, 2^{nR}\}$.
- The uniform distribution over the index set $W \in \{1, 2, \ldots, 2^{nR}\}$ induces a distribution on the input codewords, which in turn induces a distribution over the input alphabet.

Converse to the Coding Theorem for Gaussian Channels

- Consider any $(2^{nR}, n)$ code that satisfies the power constraint, that is,
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \leq P$$
for $w = 1, 2, \ldots, 2^{nR}$.
- Proceeding as in the converse for the discrete case, let W be distributed uniformly over $\{1, 2, \ldots, 2^{nR}\}$.
- The uniform distribution over the index set $W \in \{1, 2, \ldots, 2^{nR}\}$ induces a distribution on the input codewords, which in turn induces a distribution over the input alphabet.
- This specifies a joint distribution on $W \to X^n(W) \to Y^n \to \hat{W}$.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Converse to the Coding Theorem for Gaussian Channels

- To relate probability of error and mutual information, we can apply Fanos inequality to obtain

$$H(W|\hat{W}) \leq 1 + nRP_{\epsilon}^{(n)} = n\epsilon_n,$$

where $\epsilon_n \to 0$ as $P_{\epsilon}^{(n)} \to 0$. Hence,

$$\begin{aligned} nR = H(W) &= I(W;\hat{W}) + H(W|\hat{W}) \\ &\leq I(W;\hat{W}) + n\epsilon_n \\ &\leq I(X^n;Y^n) + n\epsilon_n \\ &= h(Y^n) - h(Y^n|X^n) + n\epsilon_n \\ &= h(Y^n) - h(Z^n) + n\epsilon_n \end{aligned}$$

Converse to the Coding Theorem for Gaussian Channels

-

$$\begin{aligned} nR &= h(Y^n) - h(Z^n) + n\epsilon_n \\ &\leq \sum_{i=1}^{n} h(Y_i) - h(Z^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Z_i) + n\epsilon_n \\ &= \sum_{i=1}^{n} I(X_i;Y_i) + n\epsilon_n \end{aligned}$$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Converse to the Coding Theorem for Gaussian Channels

- Here $X_i = x_i(W)$, where W is drawn according to the uniform distribution on $\{1, 2, \ldots, 2^{nR}\}$.

Converse to the Coding Theorem for Gaussian Channels

- Here $X_i = x_i(W)$, where W is drawn according to the uniform distribution on $\{1, 2, \ldots, 2^{nR}\}$.
- Now let P_i be the average power of the ith column of the codebook, that is,

$$P_i = \frac{1}{2^{nR}} \sum_{w} x_i^2(w)$$

Converse to the Coding Theorem for Gaussian Channels

- Here $X_i = x_i(W)$, where W is drawn according to the uniform distribution on $\{1, 2, \ldots, 2^{nR}\}$.
- Now let P_i be the average power of the ith column of the codebook, that is,

$$P_i = \frac{1}{2^{nR}} \sum_w x_i^2(w)$$

- Then, since $Y_i = X_i + Z_i$ and since X_i and Z_i are independent, the average power EY_i^2 of Y_i is $P_i + N$.

Converse to the Coding Theorem for Gaussian Channels

- Here $X_i = x_i(W)$, where W is drawn according to the uniform distribution on $\{1, 2, \ldots, 2^{nR}\}$.
- Now let P_i be the average power of the ith column of the codebook, that is,

$$P_i = \frac{1}{2^{nR}} \sum_w x_i^2(w)$$

- Then, since $Y_i = X_i + Z_i$ and since X_i and Z_i are independent, the average power EY_i^2 of Y_i is $P_i + N$.
- Hence, since entropy is maximized by the normal distribution,

$$h(Y_i) \leq \frac{1}{2} \log 2\pi e(P_i + N)$$

Converse to the Coding Theorem for Gaussian Channels

- Continuing with the inequalities of the converse, we obtain

$$
\begin{aligned}
nR &\leq \sum (h(Y_i) - h(Z_i)) + n\epsilon_n \\
&\leq \sum \left(\frac{1}{2} \log(2\pi e(P_i + N)) - \frac{1}{2} \log 2\pi e N \right) + n\epsilon_n \\
&= \sum \frac{1}{2} \log \left(1 + \frac{P_i}{N} \right) + n\epsilon_n
\end{aligned}
$$

Converse to the Coding Theorem for Gaussian Channels

- Continuing with the inequalities of the converse, we obtain

$$
\begin{aligned}
nR &\leq \sum (h(Y_i) - h(Z_i)) + n\epsilon_n \\
&\leq \sum \left(\frac{1}{2} \log(2\pi e(P_i + N)) - \frac{1}{2} \log 2\pi e N \right) + n\epsilon_n \\
&= \sum \frac{1}{2} \log \left(1 + \frac{P_i}{N} \right) + n\epsilon_n
\end{aligned}
$$

- Since each of the codewords satisfies the power constraint, so does their average, and hence

$$
\frac{1}{n} \sum_i P_i \leq P,
$$

Converse to the Coding Theorem for Gaussian Channels

- Since $f(x) = \frac{1}{2}\log(1+x)$ is a concave function of x, we can apply Jensens inequality to obtain

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(1+\frac{P_i}{N}\right) \leq \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{N}\right)$$
$$\leq \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

Converse to the Coding Theorem for Gaussian Channels

- Since $f(x) = \frac{1}{2}\log(1+x)$ is a concave function of x, we can apply Jensens inequality to obtain

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(1+\frac{P_i}{N}\right) \leq \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{N}\right)$$
$$\leq \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

- Thus $R \leq \frac{1}{2}\log(1+\frac{P}{N}) + \epsilon_n, \epsilon_n \rightarrow 0$, and we have the required converse.

The Gaussian Channel

- If the power constrained Gaussian channel is bandlimited to W Hz, and noise power spectral density is given by $N_0/2$, the capacity is given by

$$C = W \log\left(1 + \frac{P}{N_0 W}\right) \text{ bits per second}$$

The Gaussian Channel

- If the power constrained Gaussian channel is bandlimited to W Hz, and noise power spectral density is given by $N_0/2$, the capacity is given by

$$C = W \log\left(1 + \frac{P}{N_0 W}\right) \text{ bits per second}$$

- For Gaussian channel with infinite bandwidth ($W \to \infty$), power P and noise spectral density $N_0/2$, the capacity is given by

$$C = \frac{P}{N_0} \log_2 e \text{ bits per second}$$

Bandlimited Gaussian Channel

Proof:

- If the channel is bandlimited to W, sampling the signal at sampling rate $\frac{1}{2W}$ is sufficient to reconstruct the signal from the samples.

Bandlimited Gaussian Channel

Proof:

- If the channel is bandlimited to W, sampling the signal at sampling rate $\frac{1}{2W}$ is sufficient to reconstruct the signal from the samples.
- If the noise has power spectral density $N_0/2$ and bandwidth W, then noise power is given by $\frac{N_0}{2}2W = N_0W$ and each of the 2WT noise samples in time T has variance given by $N_0WT/2WT = N_0/2$.

Bandlimited Gaussian Channel

Proof:

- If the channel is bandlimited to W, sampling the signal at sampling rate $\frac{1}{2W}$ is sufficient to reconstruct the signal from the samples.
- If the noise has power spectral density $N_0/2$ and bandwidth W, then noise power is given by $\frac{N_0}{2}2W = N_0W$ and each of the 2WT noise samples in time T has variance given by $N_0WT/2WT = N_0/2$.
- Let the channel be used over the time interval $[0, T]$, then power per sample is given by $PT/2WT = P/2W$.

Bandlimited Gaussian Channel

Proof:

- If the channel is bandlimited to W, sampling the signal at sampling rate $\frac{1}{2W}$ is sufficient to reconstruct the signal from the samples.
- If the noise has power spectral density $N_0/2$ and bandwidth W, then noise power is given by $\frac{N_0}{2}2W = N_0W$ and each of the 2WT noise samples in time T has variance given by $N_0WT/2WT = N_0/2$.
- Let the channel be used over the time interval $[0, T]$, then power per sample is given by $PT/2WT = P/2W$.
- Noise variance per sample is given by $\frac{N_0}{2}2W\frac{T}{2WT} = N_0/2$.

Bandlimited Gaussian Channel

Proof:

- Capacity per sample is given by

$$C = \frac{1}{2}\log\left(1 + \frac{\frac{P}{2W}}{\frac{N_0}{2}}\right) = \frac{1}{2}\log\left(1 + \frac{P}{N_0 W}\right) \quad \text{bits per sample}$$

Bandlimited Gaussian Channel

Proof:

- Capacity per sample is given by

$$C = \frac{1}{2}\log\left(1 + \frac{\frac{P}{2W}}{\frac{N_0}{2}}\right) = \frac{1}{2}\log\left(1 + \frac{P}{N_0 W}\right) \quad \text{bits per sample}$$

- Since there are $2W$ samples per second, the capacity is given by

$$C = W \log\left(1 + \frac{P}{N_0 W})\right) \quad \text{bits per second}$$

Bandlimited Gaussian Channel

Proof:

- Capacity per sample is given by

$$C = \frac{1}{2}\log\left(1 + \frac{\frac{P}{2W}}{\frac{N_0}{2}}\right) = \frac{1}{2}\log\left(1 + \frac{P}{N_0 W}\right) \quad \text{bits per sample}$$

- Since there are $2W$ samples per second, the capacity is given by

$$C = W\log\left(1 + \frac{P}{N_0 W})\right) \quad \text{bits per second}$$

- For $(W \to \infty)$, we have

$$C = \frac{P}{N_0}\log_2 e \quad \text{bits per second}$$