Information Theory: Principles and Applications

Tiago T. V. Vinhoza

May 28, 2010

Differential Entropy

- Entropy of a continuous random variable.
- Let X be a random variable with cumulative distribution function $F_X(x)$ and probabliity density function $p_X(x)$.

$$h(X) = -\int_S p_X(x) \log p_X(x) dx$$

  where S is the supporting set of the random variable X, that is, the set where $p_X(x) > 0$.
- Like the discrete case, the differential entropy is only dependent of $p_X(x)$.

Differential Entropy: Example 1

- Uniform distribution.
- Consider a random variable uniformly distributed from 0 to a.

$$
\begin{aligned}
h(X) &= -\int_S p_X(x)\log p_X(x)dx \\
&= -\int_0^a \frac{1}{a}\log\frac{1}{a}dx = \log a
\end{aligned}
$$

- Note that for $a < 1$, $\log a < 0$, so the differential entropy can be negative.
- The volume of the support set, $2^{h(X)} = 2^{\log a} = a$ is always a non-negative quantity.

Differential Entropy: Example 2

- Gaussian distribution with zero mean and variance σ^2.

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

$$\begin{aligned}
h(X) &= -\int_S p_X(x) \ln p_X(x) dx \\
&= -\int p_X(x) \ln \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \right] dx \\
&= \int p_X(x) \ln \left[\sqrt{2\pi\sigma^2} \right] dx + \int x^2/2\sigma^2 p_X(x) dx \\
&= \frac{1}{2} \ln 2\pi\sigma^2 + \frac{E[X^2]}{2\sigma^2} \\
&= \frac{1}{2} \ln 2\pi\sigma^2 + \frac{1}{2} \\
&= \frac{1}{2} \ln 2\pi e\sigma^2 \text{ nats } = \frac{1}{2} \log 2\pi e\sigma^2 \text{ bits.}
\end{aligned}$$

Relations Between Differential Entropy and Discrete Entropy

- Consider a random variable X with density $p_X(x)$.
- Divide the range of X into bins of length Δ.
- Mean value theorem:

$$p_X(x_i) = \int_{i\Delta}^{(i+1)\Delta} p_X(x)dx$$

- Define the quantized random variable X^Δ

$$X^\Delta = x_i, \quad \text{if } i\Delta \leq X \leq (i+1)\Delta.$$

- $P(X^\Delta = x_i) = f(x_i)\Delta$.

Relations Between Differential Entropy and Discrete Entropy

- The entropy of this quantized variable is

$$
\begin{aligned}
H(X^{\Delta}) &= -\sum_{i=-\infty}^{\infty} p_i \log p_i \\
&= -\sum_{i=-\infty}^{\infty} p_X(x_i)\Delta \log(p_X(x_i)\Delta) \\
&= -\sum_{i=-\infty}^{\infty} p_X(x_i)\Delta \log(p_X(x_i)) - \sum_{i=-\infty}^{\infty} p_X(x_i)\Delta \log(\Delta) \\
&= -\sum_{i=-\infty}^{\infty} p_X(x_i)\Delta \log(p_X(x_i)) - \log(\Delta)
\end{aligned}
$$

Relations Between Differential Entropy and Discrete Entropy

- If the density $p_X(x)$ is Riemann integrable, then

$$H(X^{\Delta}) + \log \Delta \quad \to \quad h(X) \quad \text{as } \Delta \to 0.$$

- The entropy of and n-bit quantization of a continuous random variable X is approximately $h(X) + n$.
- Example: If X has an uniform distribution on $[0, 1]$, and we let $\Delta = 2^{-n}$, then $h(X) = 0$, $H(X^{\Delta}) = n$ and n bits suffice to describe X with n bits accuracy.
- Example: If X has an uniform distribution on $[0, 1/8)$. The first three bits after the decimal point are zero. To describe X with n bit precision, we need only $n - 3$ bits, which agrees with $h(X) = -3$.

Joint Differential Entropy

- The differential entropy of a random vector $\mathbf{X}^n$ composed of the random variables $X_1, X_2, \dots, X_n$ with density $p_{\mathbf{X}^n}(\mathbf{x}^n)$ is defined as

$$h(\mathbf{X}^n) = -\int\int \dots \int p_{\mathbf{X}^n}(\mathbf{x}^n) \log p_{\mathbf{X}^n}(\mathbf{x}^n) d\mathbf{x}^n$$

Joint Differential Entropy: Example

- Entropy of a multivariate Gaussian distribution: Let $X_1, X_2, \ldots X_n$ form a Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K}$, that is, $\mathbf{X}^n \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$.

$$h(\mathbf{X}^n) = h(X_1, X_2, \ldots, X_n) = \frac{1}{2} \log(2\pi e)^n |\mathbf{K}|.$$

where $|\mathbf{K}|$ denotes the determinant of the covariance matrix $\mathbf{K}$.

Conditional Differential Entropy

- If X, Y have a joint pdf $p_{XY}(x, y)$, the conditional differential entropy $h(X|Y)$ is defined as

$$h(X|Y) = -\int\int p_{XY}(x, y) \log p_{X|Y=y}(x) dx dy$$

- As $p_{XY}(x, y) \log p_{X|Y=y}(x) p_Y(y)$, we can write

$$h(X, Y) = h(X|Y) + h(Y)$$

Relative Entropy

- Is a measure of the distance between two continuous distributions.
- The relative entropy between two probability density functions $f_X(x)$ and $g_X(x)$ is defined as:

$$D(f_X(x)||g_X(x)) = \int f_X(x) \log \frac{f_X(x)}{g_X(x)} dx$$

Relative Entropy

- $D(f_X(x)||g_X(x)) \geq 0$ with equality if and only if $f_X(x) = g_X(x)$.
- $D(f_X(x)||g_X(x)) \neq D(g_X(x)||f_X(x))$

Mutual Information

- The mutual information of two continuous random variables X and Y is defined as the relative entropy between the joint probability density $p_{XY}(x,y)$ and the product of the marginals $p_X(x)$ and $p_Y(y)$

$$\begin{aligned} I(X;Y) &= D(p_{XY}(x,y)||p_X(x)p_Y(y)) \\ &= \int\int p_{XY}(x,y)\log\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}dxdy \end{aligned}$$

- $I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$.
- Mutual information of continuous random variables is the limit of the mutual information between their quantized versions

$$\begin{aligned} I(X^\Delta;Y^\Delta) &= H(X^\Delta) - H(X^\Delta|Y^\Delta) \\ &\approx h(X) - \log\Delta - (h(X|Y) - \log\Delta) = I(X;Y) \end{aligned}$$

Properties

- Translation does not change the differential entropy

$$h(X+c) = h(X)$$

- Multiplication by a constant

$$h(aX) = h(X) + \log|a|$$

- Same property holds for random vectors

$$h(A\mathbf{X}^n) = h(\mathbf{X}^n) + \log|A|$$

where $|A|$ is the absolute value of the determinant of A.

Properties

- The multivariate Gaussian distribution maximizes the entropy over all distribution with the same covariance matrix.

$$h(\mathbf{X}^n) \leq \frac{1}{2} \log(2\pi e)^n |\mathbf{K}|.$$

with equality if and only if $\mathbf{X}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$.

Asymptotic Equipartition Property

- Like the discrete case, we can define a typical set and characterize it
- Let $X_1, X_2, \ldots, X_n$ be a sequence of iid random variables with probability density function $p_X(x)$.

$$-\frac{1}{n} \log p_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) \rightarrow E[-\log p_X(x)] = h(X)$$

- Like the discrete case, this results follows from the weak law of large numbers

Asymptotic Equipartition Property

- The *typical set* $A_\epsilon^{(n)}$ with respect to $p_X(x)$ is the set of sequences $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ with the following property:

$$A_\epsilon^{(n)} = \left\{ \mathbf{x}^n : \left| \frac{-\log p_{\mathbf{X}^n}(\mathbf{x})}{n} - h(X) \right| \leq \epsilon \right\}$$

- The properties of the typical set for continuous random variables are the same as the ones for the discrete case.

Asymptotic Equipartition Property

- The volume of the typical set for continuous random variablesis the analog of the cardinality of the typical set for the discrete case.
- The volume $\mathrm{Vol}(A)$ of a set $A \in \mathbb{R}^n$ is defined as

$$\mathrm{Vol}(A) = \int_A dx_1 dx_2 \ldots dx_n$$

Asymptotic Equipartition Property: Properties

- $P(\mathbf{X}^n \in A_\epsilon^{(n)}) > 1 - \epsilon$ for n sufficiently large.
- $\mathrm{Vol}(A_\epsilon^{(n)}) \leq 2^{n(h(X)+\epsilon)}$ for all n.
- $\mathrm{Vol}(A_\epsilon^{(n)}) \geq (1-\epsilon)2^{n(h(X)-\epsilon)}$ for n sufficiently large.
- The results for joint typicality follows the ones for the discrete case.

Gaussian Channel

- It is a discrete-time channel where the output at time i, Y_i is the sum of the input X_i and the Gaussian noise Z_i

$$Y_i = X_i + Z_i, \qquad Z_i \sim \mathcal{N}(0, N)$$

- Noise is assumed to be independent from input.
- Noiseless case: infinite capacity. Any real number can be transmitted without error.
- Unconstrained inputs: infinite capacity. Even with noise, we can choose the inputs arbitrarily far apart, so that they are distinguishable at the output with probability of error as small as we want.

Gaussian Channel

- Limitations on the input: Power constraint
- Average power constraint is assumed. For a any length-n codeword $(x_1, x_2, \ldots x_n)$, it is required that

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

Gaussian Channel: A suboptimal use

- We want to send one bit over the channel in one use of the channel.
- Given the power constraint, we have two possibilities of signals to transmit $\sqrt{P}$ and $-\sqrt{P}$.
- From the Digital Communications class:

$$P_e = Q\left(\sqrt{\frac{P}{N}}\right)$$

- The continuous channel was converted in a discrete binary symmetric channel with crossover probability P_e.
- Discrete channels are more practical when it comes to process the output signal for error correction. However, some information is lost in the quantization.

Capacity of the Gaussian Channel

- The information capacity of the Gaussian channel with power constraint is defined as

$$C = \max_{p_X(x):E[X^2]\leq P} I(X;Y)$$

Capacity of the Gaussian Channel

- Expanding $I(X;Y)$

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

- For the noise term we have that $h(Z) = \frac{1}{2}\log 2\pi e N$.
- The variance of Y is given by

$$E[Y^2] = E[(X+Z)^2] = E[X^2] + 2E[X]E[Z] + E[Z^2] = P + N$$

- The entropy of Y is bounded by $\frac{1}{2}\log 2\pi e(P+N)$

Capacity of the Gaussian Channel

- Expanding $I(X;Y)$

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

- This maximum value is achieved if $X \sim \mathcal{N}(0,P)$, hence, the capacity of the Gaussian channel is

$$C = \max_{p_X(x):E[X^2]\leq P} I(X;Y) = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

Coding Theorem for the Gaussian Channel

- Capacity is the supremum of all achievable rates
- Arguments are similar to those for the discrete case.
- Some definitions + Achievability + Converse.

Definition: Code for a channel

- An (M, n) code for the Gaussian channel with power constraint P consists of the following:
  - An index set $\{1, 2, \dots, M\}$.
  - An encoding function $\mathbf{X}^n : \{1, 2, \dots, M\} \rightarrow \mathcal{X}^n$, that generates codewords $\mathbf{X}^n(1), \dots, \mathbf{X}^n(M)$ that satisfies the power constraint P.

$$\frac{1}{n} X_i^n(w) \leq P, \qquad w = 1, 2, \dots, M.$$

  - A decoding function

$$g : \mathcal{Y}^n \rightarrow \{1, 2, \dots, M\}.$$

which assigns a guess to each possible received vector

Definition: Rate of a code

- The rate R of an (M, n) code is:

$$R = \frac{\log M}{n} \text{ bits per transmission.}$$

- A rate R is *achievable* if there exists a sequence of $(2^{\lceil nR \rceil}, n)$ codes such that the maximal probability of error goes to zero as n goes to infinity.

Coding Theorem for the Gaussian Channel

- Consider any codeword of length n. The received vector is Gaussian distributed with mean equal to the transmitted codeword and variance equal to the noise variance.
- With high probability, the received vector is contained in a sphere of radius $\sqrt{n(N+\epsilon)}$ centered at the true codeword.
- Decoding rule: Assign every received vector that falls into the sphere to the codeword corresponding to the center of the sphere $\rightarrow$ low probability of error.

Coding Theorem for the Gaussian Channel

- How many codewords can be chosen?
- Volume of an n-dimensional sphere: $A_n r^n$, where r is the radius.
- Received vector space: sphere of radius $n(N+P)$.
- Transmitted codeword space: sphere of radius nN.
- The maximum number of non-intersecting decoding spheres in this volume is

$$\frac{A_n(n(N+P))^{(n/2)}}{A_n(nN)^{(n/2)}} = 2^{n/2\log(1+P/N)}$$

- Rate of this code $1/2\log(1+P/N)$

Coding Theorem for the Gaussian Channel: Achievability

- Generate a codebook in which all codewords satisfy the power constraint. Each element of the codeword will be generated as a Gaussian random variable with variance $P - \epsilon$. For large n, we have that $\frac{1}{n}\sum X_i^2 \rightarrow P - \epsilon$.
- Not all codewords satisfy the power constraint. They are not discarded.
- Let $X_i(w)$, $i = 1, 2, \ldots n$ and $w = 1, 2, \ldots 2^{nR}$ be iid Gaussian random variables with zero mean and variance $P - \epsilon$ forming codewords $X^n(1)$, $X^n(2)$, ..., $X^n(2^{nR}) \in \mathbb{R}^n$.
- Reveal the code to transmitter and receiver.

Coding Theorem for the Gaussian Channel: Achievability

- Message W is chosen according to a uniform distribution, that is, $P(W = w) = 2^{-nR}$, for $w = 1, 2, \ldots, 2^{nR}$.
- The codeword $\mathbf{X}^n(w)$, corresponding to the w-th row of matrix $\mathbf{C}$ is sent over the channel.
- The receiver gets sequence $\mathbf{Y}^n = \mathbf{X}^n(w) + \mathbf{Z}^n$.
- Receiver guesses message using typical set decoding
- Receiver declares that index i was sent if
  - $(\mathbf{X}^n(i), \mathbf{Y}^n)$ are jointly typical.
  - there is no other index j such that $(\mathbf{X}^n(j), \mathbf{Y}^n)$ are jointly typical.
- Otherwise the receiver declares and error.
- The receiver also declares an error if the chosen codeword does not satisfy the power constraint

Coding Theorem for the Gaussian Channel: Achievability

- Assuming WLOG that $W = 1$ was sent.

$$P(\mathcal{E}) = P(\mathcal{E}|W = 1)$$

- Defining the events

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

$$E_i = \{(\mathbf{X}^n(i), \mathbf{Y}^n) \text{ is in } A_\epsilon^{(n)}\}, \quad i = 1, 2, \ldots, 2^{nR}$$

- The error events in our case are
  - E_0 means that the codeword violates the power constraint.
  - $\overline{E_1}$, that is, the complement of E_1 occurs. This means that $\mathbf{Y}^n$ and $\mathbf{X}^n(1)$ are not jointly typical.
  - E_2 or E_3 or ... $E_{2^{nR}}$ occurs. This means that a wrong codeword is jointly typical with $\mathbf{Y}^n$.

Coding Theorem for the Gaussian Channel: Achievability

- Evaluating

$$
\begin{aligned}
P(\mathcal{E}|W=1) &= P(E_0 \cup \overline{E_1} \cup E_2 \cup E_3 \cup \ldots \cup E_{2^{nR}}) \\
&\leq P(E_0) + P(\overline{E_1}) + \sum_{i=2}^{2^{nR}} P(E_i)
\end{aligned}
$$

- By the weak law of large numbers, $P(E_0) < \epsilon$ for sufficiently large n.
- By the joint AEP, $P(\overline{E_1}) < \epsilon$ for sufficiently large n.
- As $\mathbf{X}^n(1)$ and $\mathbf{X}^n(i)$ are independent (code generation procedure), it follows that $\mathbf{Y}^n$ and $\mathbf{X}^n(i)$ are also independent if $i \neq 1$. Hence, from the joint AEP

$$
P(E_i) \leq 2^{-n(I(X;Y)-3\epsilon)} \quad \text{if } i \neq 1.
$$

Coding Theorem for the Gaussian Channel: Achievability

- Evaluating

$$
\begin{aligned}
P(\mathcal{E}|W=1) &\leq \epsilon + \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)} \\
&= 2\epsilon + (2^{nR}-1)2^{-n(I(X;Y)-3\epsilon)} \\
&\leq 2\epsilon + (2^{nR})2^{-n(I(X;Y)-3\epsilon)} \\
&= 2\epsilon + (2^{n3\epsilon})2^{-n(I(X;Y)-R)} \\
&\leq 3\epsilon
\end{aligned}
$$

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

Coding Theorem for the Gaussian Channel: Achievability

- If $R < I(X;Y)$, we can choose ϵ and n so that the average probability of error over all codebooks is less than 3ϵ.
- If the input distribution $p_X(x)$ is the one that achieves the channel capacity C, then the achievability condition is replaced by $R < C$.
- If the average probability of error over all codebooks is less than 3ϵ, than there exists at least one codebook with an average probability of error $P_e^{(n)} \leq 3\epsilon$.
- Same procedure as in the discrete case: select the best half of codewords (note that the codewords that does not satisfy the power constraint are eliminated in this step).

Coding Theorem for the Gaussian Channel: Converse

- 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

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

- Taking some steps similar to the discrete case

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

Coding Theorem for the Gaussian Channel: Converse

- In this case $X_i = X_i(W)$, where W drawn according to an uniform distribution on $\{1, 2, \ldots, 2^{nR}\}$.
- Let P_i be the average power of the $i - th$ column of the codebook.

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

- Since $Y_i = X_i + Z_i$ and since X_i and Z_i are independent, the average power of Y_i is equal to $P_i + N$. As the entropy is maximized by the Gaussian distribution

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

Coding Theorem for the Gaussian Channel: Converse

- So

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

- Dividing by n and applying Jensen's inequality we get

$$
\begin{aligned}
R &\leq \epsilon_n + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\log(1 + P_i/N) \\
&\leq \epsilon_n + \frac{1}{2}\log\left[1 + \frac{1}{n}\sum_{i=1}^{n} P_i/N\right]
\end{aligned}
$$

Coding Theorem for the Gaussian Channel: Converse

- Since each codeword satisfies the power constraint, so their average. Hence

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

$$R \leq \epsilon_n + \frac{1}{2}\log\left[1 + P/N\right]$$

- Since $\epsilon_n \to 0$ as the probability of error goes to zero, we have the required converse.

Parallel Gaussian Channels: Waterfilling

- Let us consider k independent Gaussian channels in parallel with a common power constraint.
- How to distribute the available power among the channels to maximize the capacity.
- Example: OFDM system with cyclic prefix.
- The j-channel output is given by

$$Y_j = X_j + Z_j \qquad Z_j \sim \mathcal{N}(0, N_j)$$

- Power constraint $E\left[\sum_{j=1}^{k} X_j^2\right] \leq P$

Parallel Gaussian Channels: Waterfilling

- The capacity of this channel is given by

$$C = \max_{p_{\mathbf{X}}(\mathbf{x}) : E\left[\sum_{j=1}^{k} X_j^2\right] \leq P} I(\mathbf{X};\mathbf{Y})$$

- Expanding $I(\mathbf{X};\mathbf{Y})$.

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= h(\mathbf{Y}) - h(\mathbf{Z}) \\ &\leq \sum_{i=1}^{n} h(Y_i) - h(Z_i) \\ &= \sum_{i=1}^{n} I(X_i;Y_i) \\ &\leq \sum_{i=1}^{n} \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right) \end{aligned}$$

Parallel Gaussian Channels: Waterfilling

- $P_i = E[X_i^2]$ and $\sum P_i = P$.
- Equalitiy is achieved by the input distribution $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$.
- $\mathbf{D}$ is a diagonal matrix with the powers $P_1, P_2, \ldots, P_k$.
- Optimization problem: Find the power allocation that maximizes the capacity subject to the power constraint. The Lagrangean is written as

$$\mathcal{L}(P_1, P_2, \ldots, P_k) = \sum_{i=1}^{n} \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right) + \lambda \left(\sum_{i=1}^{k} P_i\right)$$

- Differentiating with respect to P_i we have:

$$\frac{1}{2}\frac{1}{P_i + N_i} + \lambda = 0.$$

Parallel Gaussian Channels: Waterfilling

- Using the Kuhn-Tucker condition, it can be shown that the following solution maximizes the capacity

$$P_i = (\nu - N_i)^+$$

where ν is chosen in a way that $\sum_{i=1}^{k}(\nu - N_i)^+ = P$.
- Power is allocated to the better channels. Sometimes the weaker channels get no power at all.

Fading Channels

- Caused by multi-path effect: signal transmitted from a transmitter may have multiple copies traversing different paths to reach a receiver.
- The received signal should be the sum of all these multi-path signals.
- If signals are in phase, they would intensify the resultant signal; otherwise, the resultant signal is weakened due to out of phase.
- Often modelled as a random process: Rayleigh fading, Rician Fading.
- From now on, we assume that the Gaussian channel is complex valued.

$$Y_i = H_i X_i + Z_i$$

Fading Channels

- Slow fading channel: The channel gain is random but remains constant for all time, that is, $H_i = h$ for all i. This is also called the quasi-static scenario.
- Conditioned on a realization of the channel h, this is a Gaussian Channel with received signal to noise-ratio $|h|^2 P/N$.
- Suppose that the transmitter encodes the data at a rate R, but the channel realization h is such that $\log(1 + |h|^2 P/N) < R$. Then, despite the code used by the transmitter, a probability of error as small as we want cannot be assured.
- The system is said to be on outage

$$P_{out}(R) = P(\log(1 + |h|^2 P/N) < R)$$

Fading Channels

- For Rayleigh fading $H \sim \mathcal{CN}(0,1)$, so the outage probability is given by
- The system is said to be on outage

$$P_{out}(R) = 1 - \exp\left(\frac{-(2^R - 1)}{P/N}\right)$$

- The difference between the Gaussian channel and the slow fading channel is that in the former, we can send data at a positive data rate with a probability of error as small as we want.
- This cannot be done for the slow fading channel as the probablility of a deep fade is non-zero. The capacity of the slow fading channel in the strict case is zero.

Fading Channels

- Performance is given in terms of ϵ-outage capacity C_ϵ. That is the largest transmission rate R such that the outage probability $P_{out}(R) < \epsilon$.
- Solving for the case $P_{out}(R) = \epsilon$ we have:

$$C_\epsilon = \log(1 + G^{-1}(1-\epsilon)P/N)$$

where G is the complementary cumulative distribution function of $|H|^2$, that is, $G(x) = P(|H|^2 > x)$.

Next Class

- Continue the analysis of fading channels.
- Strategies to cope with fading and improve receiver performance.