Random Signals and Noise Notes

ELC5354

Spring 2025

Lecture by Professor Robert Marks, Notes taken by Chris F Lin

1 Tuesday 1/21/2024

- Probability is a model
- RNG is only pseudorandom in most computers; totally deterministic.

$$\mathbb{X}_{n+1} = frac(\mathbb{X}_n + n)^5$$

- ex. of an earlier RNG algorithm
- forecastable usable NN, not true RNG
- *quantum collapse* is claimed to be only truly random event.

2 Thursday 1/23/2025

- Last time, discussed PRNG, a recursive formula. What's really random is X_0, the initial case. can be referred to as *random seed*, so makes it random in this aspect
- when thinking of probability, we think of models: three models
  1. **Classical Definition**
     - Use to solve unfinished problems
     - $P[event] = \frac{number of desired events}{number of all events}$
     - there is a problem with this definition, can be seen in the *chord length problem*, or Bertrand's paradox. Need ot propertly define "random", but we usually don't get ambituity like this
     - another example, monty hall problem

       Exercise 1. Write a solution to the Monty Hall problem

       Exercise 2. There exists three doors, each with a different sum of money. You choose a door, and the door is opened. Decide if you want ot open the door. This is repeated if you choose to not take the money. You only get to keep the money if you pick the door with the highest amount. Dont know the amounts but do know all three are different. Come up with a strategy.
  2. **Relative Frequency**
     - experimental way

       Example 1. Estimating π by randomly throwing darts, calculating ratio of square and enscribed circle

       Example 2. Buffon's needle: $P(\text{needle crosses line}) = \frac{2a}{\pi b}$

3. **Axiomatic Definition**

   Definition 1 (Axioms)**.** S: universal set
   $A1 : P(A) >= 0$
   $A2 : P(S) = 1$
   $A3a$: if $P(A \cap B) = \emptyset$, then $P(A \cap B) = P(A) + P(B)$
   $A3b$: if disjoint A_j, then $\cup P(A_j) = \sum P(A_j)$

- We have different types of events. Typically of events in the future, but will sometimes discuss events in the past

  Definition 2 (Independent Events)**.** If A, B are independent events, then $P(AB) = P(A)P(B)$

  Definition 3 (Mutually Exclusive Events)**.** If A, B are mutually exclusive, then $A\hat{B} = \emptyset$, and so $P(AB) = 0$ and $P(A+B) = P(A) + P(B)$. For more than 2 events, have to check all three for independence.

  Definition 4 (Conditional Probability)**.** $P(A|B)$ is probability of A given B.
  $P(A|B) = \frac{P(AB)}{P(B)}$

  Theorem 1 (Theorem of Total Probability)**.** If A, B independent then $P(A|B) = P(A)$.

- for Mutually exclusive, conditional is always 0.

  Definition 5 (Partition)**.** Mutually Exclusive, then $\cup A_i = S, A_i A_j = 0 \forall i \neq j$

  Theorem 2 (Theorem of Total Probability)**.** for partitions A_i of S, $P(B) = \sum_i P(B|A_i)P(A_i)$

  Proof.

$$P(A_i|B) = P(A_i|B) = P(B|A_i)$$
$$\to P(A_i|B) = \frac{P(A|B_i)P(A_i)}{P(B)}$$
$$= \frac{P(B|A_i)P(A_i)}{\sum P(B|A_i)P(A_i)}$$

  □

- Let's look at $P(B|A)$ and how it relates to $P(A|B)$. can think of it as training and applying

  Example 3. Let B be all messages contianing the term "Nigerian Prince", A be all spam messages. Can use $P(B|A)$ to train a spam filter, then $P(A|B)$ to use it in practice.

- We call this *Bayesian Inference*, where we calculate probabilities based on results.

3 Tuesday 1/28/2025

An example of Bayes' theorem:

Example 4. Bin A has 2 white balls, 2 black balls. Bin B has 2 white balls, 1 black ball. We assume the probability of choosing from A is $\frac{1}{3}$, and the probably of choosing from B is $\frac{2}{3}$. Overall, the probability of choosing a white ball is:

$$\begin{aligned} P(\text{white ball}) &= P(\text{white}|A) \cdot P(A) + P(\text{black}|B) \cdot P(B) \\ &= \frac{1}{2} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{2}{3} \\ &= \frac{11}{18} \end{aligned}$$

Baye's asks: if you chose a black ball, what's the probability it came from bin A? Baye's theorem looks backwards:

$$\begin{aligned} P(\text{white}|A) &= P(\text{white}|A) \cdot P(A) \\ &= P(A|\text{white}) \cdot P(\text{white}) \end{aligned}$$

Therefore, rearranging, we get:

$$\begin{aligned} P(A|\text{white}) &= \frac{P(\text{white}|A) \cdot P(A)}{P(\text{white})} \\ &= \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{11}{18}} \\ &= \frac{3}{11} \end{aligned}$$

So what is a random variable? Let's say we have 2 die possibilities (here we use Bernoulli's principle of insufficient reason; assume all die outcomes are equal). Many times with events, we dont have a numerical outcome. A random variable is assigning a number to an outcome.

Example 5. if heads(H), say $X = 12$
if tails(T), say $X = 47$

For each RV, X, each event has an equal probabiltiy to happen. So what is it assigned? Introduce concept of CDF:

Definition 6 (Cumulative distribution function(CDF), $F_X(x)$)**.**

$$F_X(x) = \Pr(X \leq x)$$

Some properties:

$$0 \leq F_X(x) \leq 1$$

$$F_X(x) \leq F_X(y), \forall x \leq y$$

We can also get p between values a, b:

$$\begin{aligned} p = \Pr(a \leq X \leq b) &= F_X(b) - F_X(a) \\ &= \Pr(X \leq b) - \Pr(X \leq a) \end{aligned}$$

Also have PDF:

Definition 7 (Probability density Function(PDF), $f_X(x)$).

$$f_X(x) = \frac{d}{dx}F_X(x)$$

Properties of pdf:

$$f_X(x) \geq 0$$

This is due to F being monotonically increasing

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Can think of histograms as PDFs, normalized to unit area. call this emperical pdf. if our RVs are discrete, then they only take on integer outputs. We can always deal with discrete RVs with delta functions $\delta(t)$.

3.1 Common RV PDFs

3.1.1 Bernoulli

p: probability of success

$$\begin{aligned}\Pr(X = k) =& 1 - p, k = 0 \\ & p, k = 1\end{aligned}$$

PDF is two delta functions, at 0 and 1

3.1.2 Binomial

$\Pr(X = k)$ k times of something occurring, repeated over n bernoullie trials. k is the RV:

$$\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

3.1.3 Geometric

$$\Pr(X = k) = p(1 - p)^k$$

number of attempts before a success Some additional comments on this: Binomial expansion: $(a + b)^n = \sum_{k=0}^{n} \binom{n}{m} a^k b^{n-k}$
Combine this with definition of a geometric series and we assume the term is less than 1 to prove that the geometric series $\sum \Pr(X = k) = 1$

3.1.4 Pascal RV(negative binomial)

X is number of trials to achieve r successes

4 Tuesday 2/11/2025

Now, discussing functions of RV.
What you do with RV, reltaed to PDFs. What hpapnes when you take nonlinearity of RV,ie $-log(X)$, X uniform.
Given $f_X(x)$ pdf, let $Y = g(X)$, find $f_Y(x)$. How about new PDF of new RV?

$$f_Y(y) = \sum_{l=1}^{n} \frac{f_X(x)}{|\frac{dg(x_l)}{dx}|}$$

A special case: if $g(x)$ is strictly increasing, then because $y = g(x)$, we can say $x = g^{-1}(y)$ also strictly increasing. Additionally, we maintain inequaltiy:

$$a < b \to g(a) < g(b)$$

Example 6. let $y = ln(x)$, $x = e^y$. by defn of cdf:

$$\begin{aligned}
F_Y(y) &= \Pr(Y \leq y) \\
&= \Pr(g(X) \leq y) \\
&= \Pr(g^{-1}(g(X)) \leq g^{-1}(y) \\
&= \Pr[X \leq g^{-1}(y)] \\
&= F_x(g^{-1}(y))
\end{aligned}$$

slightly more difficult to get pdf, need to use calculus and chain rule:

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d}{dy} F_X(g^{-1}(y)) \\
&= \frac{dg^{-1}(y)}{dy} f_X(g^{-1}(y)) \\
&= \frac{f_X(x)}{\frac{dg(x)}{dx}}
\end{aligned}$$

For a smooth function, we can break into strictly increasing or strictly decreasing. We need to ensure: $f_Y(y)dy = f_X(x)dx$
Let's examine $f_y(x) = f_x(x)\frac{dx}{dy} = \frac{f_X(x)}{\frac{dy}{dx}}$. the second part is the same as what we calculated before. This appeals to the conservation of probability mass. we now have two derivations. this gives way to think of nonlinearities. Some simple transformations:

- Transposition: $Y = -X, f_y(y) = f_X(-y)$
- Shift: $Y = X + S, f_Y(y) = f_X(y - S)$
- Scale: $Y = aX, f_Y(y) = \frac{1}{|a|} f_X(\frac{y}{a})$. stretches horizontally, so height will change to maintain constant area
- Modulus: $Y = |X|, f_Y(y) = (f_X(y) + f_X(-y))u(y)$; this is a double mapping so uses step function $u(x)$
- Shift + Scale: $Y = aX + S, f_Y(y) = \frac{1}{|a|} f_X(\frac{y-S}{a})$

what happens if you have a discontinuity? we map over to a delta function, so identically zero in that interval.Remember that we're doing continuous and not discrete right now, so the prob of a specific event occurring is 0.

5 Thursday 2/13/2025

So far, have done analysis: given $X, y = g(x)$, what to find Y, $f_Y(y)$. What if we have f_x, f_y, can we find $g(x)$? we note that if we can, $g(x)$ is not unique. Sometimes we want to constrain $g(x)$. one example is if $g(x) = F_Y^{-1}(x)$. so if X is uniform, one $g(x)$ is always $f_Y(y)$.
If $f_Y(y) = e^{-Y}U(y)$, $F_Y(y) = (1 - e^{-}y)U(x)$, then $x = -ln(1 - y)$.
So what if we take an arbitrary pdf, map it to $U[0, 1]$?

Moving on, go into moments. In general, we have: $E[g(x)] = \int g(x)f_X(x)dx$. Averages converge to expected values by law of large numbers. Two of the most common measures:

- Mean: $E(x) = \int xf_X(x)dx$.
  Measure of central tendency, Median is another measure of central tendency, mode is another. Mean typically used because it has most analytic tractability
- Dispersion/Variance: $\sigma^2 = E[x^2] - E[x]^2$
  What is the units? same as mean (ie if volts, then also volts).
  Range is another measure of dispersion. If we do $E(x^2) + E(x)^2$, then estimate the sample variance.
- Other terms: $E(x^m)$ is m^{th} moment

6 2/20/2025 Thursday

Some more review on expected values:

$$E(g(x)) = \int g(x)f_X(x)dx$$

nth moment:

$$E(x^n) = \int x^n f_X(x)dx$$

The mean is 1st moment; variance uses both first and second moments. Markov and Chebychev inequalities show that the variance represents how spread out the values are.
Some RVs have mean infinite:

Example 7.

$$F_X(x) = 1 - \frac{1}{x}, x > 1$$

Example 8. Cauchy RV: 2nd moment is ∞

$$\frac{x}{x^2+1} \to 1$$

Example 9 (St Petersburg Gambling Problem)**.** N consecutive heads and 1 tail, you get 2^N dollars, how much should you bet?

$$\sum_N P(X=k)2^k - c = \sum_N 1 - c = \infty$$

Therefore, we should be willing to bet any amount. In reality, we have a finite nubmer of games, number of time.

A note, your outcome may not be realized. for example, a coinflip is 1/2, but you cannot get .5, only 0 or 1. Now we look at fourier transform and applications in probability via characteristic functions of RV.

Definition 8 (Characteristic Function $\Phi_X(\omega)$)**.**

$$\Phi_X(\omega) = E(e^{j\omega x}) = \int e^{j\omega x} f_X(x)dx$$

Some properties:

$$|\Phi_X(\omega)| \leq \Phi_W(0) = 1$$

This is due to directly substituting 0 into the equation, and f_X has integral 1.

$$E(X^n) = (-j)^n \frac{d^n \Phi_X(\omega)}{dw^n}|_{\omega=0}$$

If we take the n^{th} derivative:

$$\begin{aligned}(\frac{x}{d\omega})^n \Phi(\omega) &= \int f_X(x)((\frac{d}{d\omega})^n e^{j\omega x})dx \\ &= \int (jx)^n f_X(x) e^{j\omega x} dx\end{aligned}$$

so evaluating at $\omega = 0$ gives us the n^{th} moment, scaled by $(-j)^n$
This is easier than calculating the moments directly if we know the FT easily.
We can also use the characteristic functions to add two PDFs together:

$$\Phi_{X+Y}(\omega) = \Phi_X(\omega)\Phi_Y(\omega)$$

An example is rolling two die and summing the results. This leads to the proof of the Central limit theorem (CLT): convolving enough distributions results in a guassian distribution; this is why we use gaussians everywhere

Definition 9 (2nd Characteristic Function $\Psi_X(x)$)**.**

$$\Psi_X(x) = ln(\Phi_X(x))$$

This allows us to get variance right away instead of calculating the moments and subtracting!

Example 10.

$$f_X(x) = (1-p)\delta(x) + p\delta(x-1)$$

Using sifting property of $\delta(x)$ to compute the characteristic function of a bernoulli trial ($\int g(x)\delta(x-\xi)dx = g(\xi)$):

$$\begin{aligned}\int f_X(x)e^{j\omega x}dx &= (1-p)e^{j\omega 0} + p \cdot e^{j\omega \cdot 1} \\ &= 1 - p + p \cdot e^{jw} = \Phi_X(\omega)\end{aligned}$$

note: We can do something similar for Z and Laplacian transforms.
For Binomials, we're adding up n repeated bernoulli trials, so this is same as convolving many bernoullis together, therefore:

$$\begin{aligned}\Phi_X(\omega) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{j\omega k} \\ &= \sum_{k}^{n} \binom{n}{k} (pe^{jk})^k (1-p)^{n-k} \\ &= (1 - p + pe^{j\omega})^n\end{aligned}$$

the last step by the binomial expansion.

We now have multiple ways we can define a RV:

- integral/derivative between f_X, F_X
- fourier transform between F_X, characteristic function Φ_X
- natural log between first and second characteristic functions Φ_X, Ψ_X

If we are workign with discrete RVs, we instead take the z-transform (pmf); both can be used to obtain mean and variance.

6.1 Some Common Models and their Characteristic functions

6.1.1 Geometric

$$\begin{aligned}\Phi_X(\omega) &= \sum p(1-p)^k e^{j\omega k} && (1)\\ &= p \sum ((1-p)e^{j\omega})^k && (2)\\ &= \frac{pe^{j\omega}}{1-(1-p)e^{j\omega}} && (3)\end{aligned}$$

6.1.2 Poisson

$$Pr(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

$$\begin{aligned}\Phi_X(\omega) &= \sum \frac{\lambda^k}{k!}e^{-\lambda}e^{\omega jk} \\ &= e^{-\lambda}\sum \frac{(\lambda e^{j\omega})^k}{k!} \\ &= e^{-\lambda}e^{\lambda e^{j\omega}} = e^{\lambda(e^{j\omega}-1)}\end{aligned}$$

6.1.3 Exponential RV

$$f_X(x) = \lambda e^{\lambda x}U(x)$$

So by direct integration,

$$\Phi_X(\omega) = \frac{\lambda}{\lambda - j\omega}$$

The Laplace RV is simply a generalization of the exponential RV

7 2/25/2025 Tuesday

note: Fourier transform is the only transform with a physical interpretation. Let's take the binomial:

$$Pr(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$$

then:

$$E(X) = \bar{X} = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$$

Tedious evaluation, so instead do characteristic function and 1st moment:

$$\Phi_X(\omega) = (1 - p + pe^{j\omega})^n$$

the second characteristic function is even easier:

$$\Psi_X(\omega) = ln\Phi_X(\omega) = ln((1 - p + pe^{j\omega})^n)$$

take derivative with respect to ω and set $\omega = 0$, get $jE(x) = jnp$.
Also:

$$\begin{aligned}\frac{d^2}{d\omega^2}\Psi_X(0) &= -var(X) \\ &= -np(1-p)\end{aligned}$$

Remember we said σ must be same units as M. however, for poisson: $\mu = \lambda$, $var = \lambda \rightarrow \sigma = \sqrt{\lambda}$, so units is $\sqrt{\dots}$. Is this a paradox? no, discrete is the reasoning: "3 counts" has no unit of measurement, technically unitless.

Theorem 3 (Markov Inequality)**.** Assume pdf: $x \geq 0$. If $f_X(x) = f_X(x)\mu(x)$, then:

$$Pr(x \geq a) \leq \frac{\bar{X}}{a} \forall a > 0$$

Proof.

$$E(x) = [\int_{x=0}^{a} + \int_{a}^{\infty}] x f_X(x) dx$$

both nonnegative, so:

$$\rightarrow \geq \int_{a}^{\infty} x f_X(x) dx \geq \int_{a}^{\infty} a f_X(x) dx = a \cdot Pr(x \geq a)$$

and result directly follows rearranging. □

Example 11.

$$f_X(x) = e^{-x}$$

we know $E(x) = \bar{X} = 1$, so

$$Pr(X \geq a) = \int_{a}^{\infty} e^{-x} dx = e^{-x}|_{a}^{-\infty} = e^{-a}$$

therefore, $\bar{X} = 1 \geq e^{-a}, \forall a > 0$.

Theorem 4 (Chebyshev Inequality)**.** $Pr(|X - \bar{X}| > a) \leq \frac{var(X)}{a^2}$

This shows us that the standard deviation is how mcuh the PDF is spread out: as variance increases, it spread out more

Proof.

$$\begin{aligned}
\mathrm{Var}(x) &= \int_{-\infty}^{\infty} (x - \bar{x})^2 f_X(x)\, dx \\
&\geq \int_{-\infty}^{\bar{x}-a} (x - \bar{x})^2 f_X(x)\, dx + \int_{\bar{x}+a}^{\infty} (x - \bar{x})^2 f_X(x)\, dx \\
&\geq \int_{|x-\bar{x}| \geq a} (x - \bar{x})^2 f_X(x)\, dx \\
&\geq a^2 \int_{|x-\bar{x}| \geq a} f_X(x)\, dx \\
&= a^2 \cdot \Pr(|x - \bar{x}| \geq a)
\end{aligned}$$

□

We also have the Chernoff bound; this uses the laplace transform. so chevyshev inequality illustrates variance as well as measures dispersion.

8 3/4/2025 Tuesday

now expanding to multidimensional RV

8.1 2D RVs

X, Y joint RVs, typically related, described by joint cdf:

$$F_{XY}(x, y) = Pr(X \leq x, Y \leq y)$$

Properties of joint cdf:

$$0 \leq F_{XY}(x, y) \leq 1$$

$$F_{XY}(\infty, \infty) = Pr(X \leq \infty, Y \leq \infty) = 1$$

$$F_{XY}(-\infty, -\infty) = Pr(X \leq -\infty, Y \leq -\infty) = 0$$

infinite number of paths to either infinite. To get 1D case:

$$\begin{aligned} F_{XY}(x, \infty) &= Pr(X \leq x, Y \leq y) && (4) \\ &= Pr(X \leq x) && (5) \\ &= F_X(x) && (6) \end{aligned}$$

Taking slice of F_{XY} will get the CDF $F_X(x)$. we call this the **marginal CDF** if $F_{XY}(x, y)$ is the **joint cdf**. Let's add the assumption that X and Y are independent. Then,

$$Pr(X \leq x, Y \leq y) = Pr(X \leq x)Pr(Y \leq y) = F_X(x)F_Y(y)$$

When we have $\Delta x, \Delta y > 0$,

$$F_{XY}(x + \Delta x, y + \Delta y) \geq F_{XY}(x, y)$$

For the 2D pdf, it is a partial deriative in x, then partial in y. Conversely, if given pdf, it is integral up to x and up to y:

$$f_{XY}(x, y) = \frac{d^2}{dxdy} F_{XY}(x, y)$$

$$F_{XY}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(\xi, \nu) d\xi d\nu$$

Some properties:

- $f_{XY}(x, y) \geq 0$
- total volume of PDF is 1: set x,y to infinite
- if independent X, Y, then F separable:

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

- define $f_{XY}(x, y)$ as the **joint PDF**, $f_X(x)$, $f_Y(y)$ are **marginal PDFs**

- can think of a normalized histogram as an emprical PDF; each row/column of a table also parginal pdfs

Note: for convergence due to rule of large numbers, we need to be careful with type of convergence. Interpretting the PDF:

$$Pr((X,Y) \in A) = \int\int_A f_{XY}(x,y)dxdy$$

To get the marginals, we integrate over one of the variables!

Example 12.

$$Pr(X > Y) = \int\int f_{XY}(x,y)dxdy \qquad\qquad = \int\int_{x=y}^{\infty} \dots dxdy$$

Example 13.

$$Pr(min(X,Y) > z) = \int_z^{\infty}\int_z^{\infty} f_X f_Y dxdy$$

Looking back at discrete and the use of dirac deltas: will utilize shifts and multiply by the associated probabilities.

$$P_{XY}(x_j, y_k) = Pr(X = x_j, Y = j_k) = \sum_j\sum_k P_{XY}(x_j, y_k)\delta(x - x_j)\delta(y - y_k)$$

Example 14. Let X be the side length of a cube, Y cube volume. We can derive everything from the CDF:

9 Thursday 3/6/2025

Definition 10 (Multidimensional Random Variables)**.** For MD-RV, $\vec{x} = (x_1, \dots x_N) \in \mathbb{R}^N$, have corresponding CDF and PDF:

$$\textbf{CDF:} F_{\vec{X}}(\vec{x}) = F_X(x_1, \dots, x_N) = \Pr(\vec{x} \leq \vec{X}) = \Pr(X_1 \leq x_1, \dots X_n \leq x_n)$$

$$\textbf{PDF:} \frac{d^n}{dx_1 \dots dx_N} F_{\vec{X}}(\vec{x})$$

We also have marginal CDFs: simply set the variables to ∞ if you do not need them, results in a lower dimension probabilty space.

$$F_{X_1,X_2}(x_1, x_2) = F_{vecX}(x_1, x_2, \infty, \dots, \infty)$$

Can do something similar for marginal PDFs:

$$f_{X_1,X_2}(x_1, x_2) = \int_{x_3} \cdots \int_{x_n} f_{X_1 \dots X_n} dx_3 \dots dx_n$$

For conditionals:

$$f_{\vec{X}}(x_3, \dots x_n | x_1, x_2) = \frac{x_{\vec{X}}(x_1, \dots x_n)}{f_{\vec{X}}(x_1, x_2)}$$

If we have independence for all dimensions, then:

$$f_{\vec{X}}(\vec{x}) = \prod_{k=1}^{n} f_{X_k}(x_k)$$

9.1 Expectation of MDim RV

Definition 11. Expectation of MDim function $g(\vec{X})$:

$$E[g(\vec{X})] = \int_{\vec{x}} g(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x}$$

like all cases, we get separability if there is independence between all X_k's:

$$E[\prod g_k(X_k)] = \prod E[g_k(X_k)]$$

We now introduce the joint characteristic function Φ by taking the higher dimension Fourier transform of the pdf:

Definition 12. Joint characteristic function $\Phi_{\vec{X}}(\vec{\omega})$:

$$\Phi_{\vec{X}}(\vec{\omega}) = \int f_{\vec{X}}(\vec{x}) e^{j\vec{\omega}^T \vec{x}} d\vec{x}$$

We just like how in 1D, we used the characteristif function to represent the sum of RVs, we do the same here.

Example 15. Say we have X, Y independent RVs, let $Z = X + Y$. then:

$$\Phi_Z(\omega) = E[e^{j\omega z}] = E[e^{j\omega(X+Y)}] = E[e^{j\omega x}]E[e^{j\omega y}] = \Phi_X(\omega)\Phi_Y(\omega)$$

can observe that both PDF and char functions are separable if we have independence.
Note that multiplication in Fourier domain is convolution in spatial. therefore, we can represent the density function of Z as:

$$f_Z(z) = f_X(z) * f_Y(z) = \int_{\tau \in \mathbb{R}} f_X(\tau) f_Y(z - \tau) d\tau$$

once again, can use sum of die outcomes as an example.
Now, let's introduce stricter conditions: **independent and identically distributed** Let's look at the sum $S = \sum X_k$. we're convolving X_k with itself n number of times, therefore our sum's characteristic function is the product of the corresponding char funcs, which are all the same due to being identically distributed. therefore:

$$\Phi_S(\omega) = \Phi_X^n(\omega)$$

Some examples of adding multiple IIDs: "if x_k is ___, then S is ___"

- gaussian $\to$ gaussian
- poisson $\to$ poisson
- binomial $\to$ binomial
- gamm $\to$ gamma
- cauchy $\to$ cauchy
- neg binomial $\to$ neg binomial

Let's take a closer look at the cauchy function. The char. func is $\Phi_X(x) = e^{-\alpha\|\omega\|}$ The derivative is undefined at 0. Second derivative is similar to original char. function, except a delta function at 0.

9.2 functions of N-D RV

We can also do functions of N-dimensional RV. have 2 types:

- 1D output: $z = g(\vec{x}) = \mathbb{R}^N \to \mathbb{R}$
- ND output: $z_k = g_k(\vec{X}); 1 \le k \le n$

introduce Leibnitze Rule to help later:

Definition 13 (Leibnitze Rule)**.**

$$\frac{d}{dz}\int_{l(z)}^{u(z)} h(x,z)dx = \frac{duz}{dz}h(u(z),z) - \frac{dl(z)}{dz}h(l(z),z) + \int_l (z)^h(z)\frac{d}{dz}h(x,z)dx$$

we'll apply this to one function of several RVs: $z = g(\vec{X})$ the cdf of z:

$$F_Z(\vec{z}) = \Pr(Z \le z) = \Pr(g(\vec{x}) \le z)$$

The challenging aspect is "finding a region":

$$R_z = \{\vec{x} | g(\vec{x}) \le z\}$$

Have to integrate over R_Z to get the CDF of Z

Example 16. let $Z = g(X,Y) = X + Y$, define $R_Z = \{(x,y)|y \le z - x\}$. Therefore:

$$F_z(z) = \Pr(Z \le z) = \Pr(g(\vec{x}) \le z) = \int_{x\in\mathbb{R}} \int_{y=\infty}^{z-x} f_{XY}(x,y)dxdy$$

If X, Y are indep, then we get a convolution. applying leibnitz:

$$f_Z(z) = \frac{d}{dz}F(z) = \frac{d}{dz}\int\int_{-\infty}^{z-x} \cdots = \int_{\mathbb{R}} F_{XY}(x, z-x)dx$$

this is a general result. Assuming X, Y are independent, $Z = X + Y$, then

$$f_X(x)f_Y(z-x) = f_Z = \int f_X(x)f_Y(z-x)dx = f_X(z) * f_Y(z)$$

This is what we saw with characteristic functions!

10 ...MISSING NOTES...

11 Tuesday 3/18/2025

Last time, we had the markov and chebychev inequalities:

Definition 14 (Markov Ineq)**.** if $f_X(x) = f_X(x)\mu(x)$, then $\Pr[x \geq z] \leq \frac{\bar{x}}{a}$

Definition 15 (Chebychev Ineq)**.** ... $\Pr[|x - \bar{x}| > a] \leq \frac{var(x)}{a^2}$

Now, let's work up ot ND RV and characterizing stochastic processes.

Example 17. Product of 2 RVs: $Z = XY$, assume $X, Y > 0$. then:

$$F_Z(z) = \Pr[Z \leq z] = \Pr[XY \leq z] = \Pr[Y \leq \frac{Z}{X}]$$

the area is everything below the curve: $y = z/x$. therefore, by def,

$$= \int_0^\infty \int_{y=0}^{z/x} f_{XY}(x, y) dy dx$$

take the derivative to get the PDF; applying leibnitz, we get

$$f_Z(z) = \frac{d}{dz} F(z) = \int_{R^+} f_{XY}(x, z/x) dx$$

If we extend this example and let X, Y be I.I.D and Uniform on $(0, 1)$, we know that $f_Z(z) = \int_0^\infty \frac{1}{x} f_{XY}(x, z/x) dx = 1$. if $x, y \in (0, 1)$. thus,

$$f_Z(z) = \int_0^1 \frac{1}{x} dx = -ln(z); 0 < z < 1$$

We observe the density function approaches 0 as we approach the origin. Look at the slides for $Z = X/Y$ example.

Now let's look at discrete expectations. If X, Y RV, can use prob mass function. 2 ways to find $E[Z]$, $z = g(\bar{X})$.

$$E[z] = E[g(Z)] = \int g(x) f_X(x) dx$$

We can first find f_Z then compute the expectations, but this requires more work.

Definition 16. Joint moments of X, Y:

$$E[X^l Y^k] = \int \int x^l y^k f_{XY}(x, y) dx dy = \sum_i \sum_n d_i^l y_n^k p_{XY}(x_i, y_n)$$

The main interesting results are when $k, l = 1, 2$. The correlation between X, Y: $E[XY]$. 0 if no correlation or X, Y are orthogonal. looking at covariance $cov(X, Y)$:

$$\text{cov}(x, y) = E[(x - \bar{x})(y - \bar{y})]$$

correlation coeff $\rho(x, y)$:

$$\rho(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y} = \frac{E(XY) - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

With enough data, we can measure these two without knowing the PDF. $\sigma^2 = E[(x - \bar{x})^2] = E[x^2] - E[x]^2$, similar to correlation coefficient definition.
In 1D, gaussian can be accurately described by two variables $(\bar{x}, \sigma)$. in ND, not so direct, but for 2D, we only need 5 variables. We can show that if we have no correlation (ie $\rho_X Y = 0$), then we get separability. **This gaussian is the only distribution where correlation and independence are the same thing**. Correlation is a weaker statement than independence.

11.1 Contours

If $g(x, y)$ has contours, then $f(g(x, y))$ has the same contours; it's just that the contour values change: if $g(x, y) = a$, then $f(g(x, y)) = f(a)$. so we can look at contours of joint gaussians. Contours of gaussian same as contours of exponential: an ellipse, 5 parameters.

12 Thurs 3/21/2025

Sums of RV: $S = \sum X_k$, mean is $\bar{S} = \frac{1}{n} \sum \bar{X}_k$. if we also assume i.i.d, then $\bar{S} = n\bar{X}$. so how about variance?

$$\begin{aligned}
\text{var}(s) &= \mathbb{E}\left[(s - \bar{s})^2\right] \\
&= \mathbb{E}\left[\left(\sum_k x_k - \sum_k \bar{x}_k\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_k (x_k - \bar{x}_k)\right)^2\right] \\
&= \mathbb{E}\left[\sum_k \sum_l (x_k - \bar{x}_k)(x_l - \bar{x}_l)\right] \\
&= \sum_k \sum_l \mathbb{E}\left[(x_k - \bar{x}_k)(x_l - \bar{x}_l)\right] \\
&= \sum_k \sum_l \text{cov}(x_k, x_l)
\end{aligned}$$

the covariance matrix is a toeplitz matrix, where each component is the covariance between the RVs: $C_{ij} = cov(X_i, Y_j)$. Adding up all elements of covariance matrix is equal to the variance of

sum of these RVs. when we do this, the RVs need not be independent.
Assume independence: then

$$cov(X_1, X_2) = E[(X_1 - \bar{X_1})(X_2 - \bar{X_2})] = E[X_1 - \bar{X_1}]E[X_2 - \bar{X_2}] = 0$$

Therefore, with independence, the covariance matrix is a diagonal matrix, made up of of the variance of the individual RVs, and $var(S) = \sum \sigma_k^2$
This makes sense because our std. dev is how uncertain we are, and uncertainty is just adding up all the uncertainties of all the other variables.
Now add IID: all the same so $var(S) = n\sigma_X^2$.
For characteristic functions:

- First: $\Phi_S(\omega) = (\Phi_X(\omega))^n$
- Second: $\Psi_S(\omega) = n\Psi_X(\omega)$

looking at examples, if we have all cauchy, then : $\Phi_X(\omega) = e^{-a|\omega|} \rightarrow \Phi_S(\omega) = (e^{-a|\omega|})^n = e^{-an|\omega|}$, another cauchy RV with parameter $a \cdot n$. If we have sum of gaussians, then we get a mean of $n\bar{X}$ and variance of $n\sigma^2$.
Remember we care often about average and not just the sum. Since we can measure the average and not the real mean. Let $A = M_n = \frac{1}{n}\sum X_k = \frac{1}{n}S$. then the characteristic function of the average is:

$$\Phi_A(\omega) = \Phi_S(\frac{\omega}{n})$$

The second characteristic function is: $\Psi_A(\omega) = n\Psi_A(\frac{\omega}{n})$. we can rederive using the derivative of the 2nd characteristic function.

$$\frac{d}{d\omega}\Psi_A(\omega) = n\frac{d}{d\omega}\Psi_X(\frac{\omega}{n}) = \Psi_X'(\frac{\omega}{n}) = j \cdot (\text{MEAN BY PREV DEFN})$$

or:

$$j\bar{A} = \Psi_A'(0) = \Psi_X'(0) = j\bar{X}$$

repeat for getting the variance:

$$\frac{d^2}{d\omega^2}\Psi_A(\omega) = n\frac{d^2}{d\omega^2}\Psi_X(\frac{\omega}{n}) = \frac{1}{n}\Psi_X''(\frac{\omega}{n})$$

as n increases, variance decreases, get more certainty. unfortunately, $var(A) = \frac{1}{n}var(X)$; smaller reduction in uncertainty. by law of large numbers, average approaches mean for large n. we can apply chebychev to see this:

$$Pr[|A - \bar{A}| < \epsilon] \geq 1 - \frac{var(A)}{\epsilon^2} = 1 - \frac{var(X)}{n\epsilon^2}$$

Definition 17. Weak law of large numbers:

$$lim_{n\rightarrow\infty} Pr[|A - \bar{X}| < \epsilon] = 1$$

can have a stronger statement:

Definition 18. Strong law of large numbers:

$$Pr[lim_{n\rightarrow\infty} A = \bar{X}] = 1$$

this is a different type of convergence

13 Tuesday 3/25/2025

last time, discussed weak/strong law of large numbers. So if you're in a game, how much risk do you take everytime if you don't ever want to hit 0? will apply law of large nubmers

Example 18. scenario is bernoulli trial with $p > \frac{1}{2}$. Start with \$D, the potential return for a win in a single trial is the bet amount.
What is the optimal bet % of your current balance for steady winnings? we want to maximize winnings. example of a scenario:
win: $D(1) = D + \%D = (1 + \%)D$, % percent bet
lose: $D(2) = D(1) - \%D(1) = (1 - \%)(1 + \%)D = (1 - \%^1)D$
lose: $D(3) = D(2) - \%D(2) = (1 + \%)(1 - \%)^2$
W wins, L loses: $(1 + \%)^W(1 - \%)LD$
So what percentage maximizes $D(n)$ for large n? Same as maximizing $B(n) = \frac{1}{n}log(D(n)) \rightarrow \frac{k}{n}log(1 + \%) + \frac{n-1}{n}log(1 - \%) + log(D)$.
as $n \rightarrow \infty$, $\frac{k}{n}$ represents chacne of winning, same/opposite with $\frac{n-k}{n}$, percent chance of losing. taking the derivative of $B(n)$, we get:

$$\frac{dB(n)}{d\%} = 0 = \frac{p}{1 + \%} - \frac{1 - D}{1 - \%} \rightarrow \% = 2p - 1$$

So if $p \leq 1/2$, bet 0%. optimal return is then:

$$D(n) = (1 + \%)^k(1 - \%)^{n-k}D = 2^n p^k(1 - p)^{n-k}$$

Can use this general formula to determine how many bets we need to get X return, eg double return or triple return, by dividing both sides by D and setting equal to target goal.

Definition 19. Stochastic Resonance is when the addition of noise improves performance

For example, say our detector has limited capabilities, assume 0 or 1 due to thresholding. using law of large numbers, can get image very similar to the original. noise is not always bad! sometimes signals can be modeled as a stochastic process.

14 Thursday 3/27/2025

last time, proved law of large numbers for chebyshev's inequality. for more nad more samples, the PDF squeezes to a delta function, closer to the mean $\bar{X}$.
Law of large numbers says if $\bar{y} = e^X$, then $E[e^X] = \bar{Y}$. Can apply this to any function of a RV. now we transition into convolutions and the central limit theorem (CLT).

14.1 Central Limit Theorem

Take nonnegative functions $g_1(x), g_2(x), g_3(x), \ldots,$ and convolve them. If we convolve enough $g_i(x)$ together, we will get a gaussian distribution! We have seen that when adding RVs, we are convolving their PDFs. Therefore, when we are averaging n i.i.d.'s, the sum is the autoconvolution n times. Below is proof of CLT

Central Limit Theorem. Let S be the sum of n RVs, i.i.d. Define

$$Z = \frac{S - \bar{S}}{stdev(S)} = \frac{S - n\bar{X}}{\sigma\sqrt{n}}$$

This is normalizing to 0 mean, standard deviation of 1 (both easy to show directly by defn as $n \to \infty$). Note that for both, we are deviding by the σ of the original RVs. Therefore, a necessary condition is that the variance is defined; this is important because it is not defined for all, most notably the **cauchy RV** (will result in a cauchy). □

Now, recall that $erf(z) = 1/2(1 - 2Q(z))$. We are going to show that limit of $z \to \infty$ approaches $1/2erf(z)$, so cdf is a gaussian.

Example 19. let X $uniform(-1/2, 1/2)$ be a rounding error. S is total rounding error. $E(X) = \$0$, $var(X) = \$1/12$. Adding RVs, the new mean is sum of means: S $N(0, \sqrt{\frac{n}{12}})$.
What's probability error more than \$10 for 10 samples? what's n when $Pr[|S| > \$100] = 1/2$?

Example 20. Bernoulli trials: we know norm is np, var is $np(1-p)$. CLT says if we do enough trials, we get a gaussian.**Demoivre-Laplace Theorem** is a special case of CLT on bernoulli trials/binomials.
Rembmer, we said for finite n, distributions typically become the same distributions with the variables changed. for infinite, they become gaussian (except cauchy). This is why we need the existence of $\bar{X}, \sigma_X$
Remember that for convolution, we need independence. we can even relax the independence criteria, but CLT is only true under certain conditions:

a usefule lemma:

$$lim_{n\to\infty}(1 - \frac{\beta}{n})^n = e^{-\beta}$$

15 Tuesday 4/1/2025

an off note: the power curve. a special case of pareto RV, "ubiquity" (eg. size of cities and nubmer of cities of that size)
Now, last time we discussed the CLT. We get a gaussian RV, assuming we adding RVs with defined mean and stdev. the Laplace-demoivre, a special case of the CLT on bernoulli trials. we also know only the cauchy RVs dont become gaussian due to undefined variables. now for the proof of the CLT

CLT Proof. Begin with LD lemma:

$$log((1 - \frac{\beta}{n})^n) = nlog(1 - \frac{\beta}{n})$$

Using l'hopitals, we get:

$$lim_{n\to\infty}log(\dots) = -\beta \to lim_{n\to\infty}(\dots) = e^{-\beta}$$

For small ϵ, $e^{\epsilon}1 + 1/\epsilon + 1/\epsilon^2 + \dots$ will now show:

$$\Phi_Z(\omega) = E(e^{j\omega Z}) \to e^{-\omega^2/2}, n \to \infty$$

To get the characteristic function, we have:

$$\begin{aligned} e^{j\omega Z} &= e^{j\omega \frac{S - n\bar{X}}{\sigma\sqrt{n}}} \\ &= e^{j\omega \frac{\sum x_k - \bar{x}}{\sigma\sqrt{n}}} \\ &= \prod_{k=1}^{n} e^{\cdots} \end{aligned}$$

therefore, $\Phi_Z(\omega) = E(e^{j\omega z}) = E(\prod_{k=1}^{n} e^{j\omega \frac{x_k - \bar{x}}{\sigma\sqrt{n}}})$ Since X_k's are iid:

$$Z(\omega) = \prod_{k=1}^{n} E(e^{j\omega()}) = (E(e^{j\omega}))^n$$

expanding the exponential term with a taylor series:

$$\Phi_Z(\omega) = E(1 + j\omega(\frac{X_k - \bar{X}}{\sigma\sqrt{n}}) - \omega^2/2(\frac{X_k - \bar{X}}{\sigma\sqrt{n}}))^n$$

the middle term is 0 due to $E() = 0$ and the right side is the definition of σ^2. note that we only need to expand to 2nd order of taylor series due to the higher terms leading to $\frac{1}{n}$, so the above becomes:

$$= (1 - \frac{\omega^2}{2}(\frac{sigma}{\sigma\sqrt{n}})^2)^n$$

using the lemma prior, we get, if we take the above for $n \to \infty$,

$$lim_{n\to\infty}(\dots) = e^{\omega^2/2}$$

□

so we know that the sum becomes gaussian, this means that the mean also becomes a gaussian with enough samples.

15.1 Confidence Intervals

$$Pr(-z \le \frac{A - \bar{X}}{\sigma/\sqrt{n}}) = 2erf(z) = p$$

z is the number of standard deviations. "How close is the average to the mean?" this is the confidence interval. may not know the standard deviation itself.

16 Tuesday 4/3/2025

Example 21. Say we know 1 vial is bad; either all pills in a vial are good or all are bad: 1g vs 1.1g (good vs bad). will take i pills from the ith container, so we only need 1 weigh to determine the bad vail.
How about if more than 1 vial is bad? can solve this still with only 1 weight: take 1, 2, 4, ... powers of 2. This leads into information theory

16.1 Shannon Information Axioms

- Small probability events have more information than large probabilty ones. information should "add" if disjoint. $log_2(p)$ is the only function that matches these axioms, though technically can be any base. base determines number of bits.
- An application is interval halving: need 4 bits to guess a square. using binary search to determine.
- for bernoulli trials, information from success is $-log_2(p)$, $-log_2(1-p)$ for failures.
- Shannon cared about average amount of information: introducing term entropy

Definition 20 (Entropy H)**.** $H = -\sum_n p_n log_2(p_n)$ is the average information

looking at the binomial entropy function $H = -plog_2(p) - (1-p)log_2(1-p)$, weget the most information when $p = 1/2$. no information if it's alwys heads or always tails.

Example 22. entropy of uniform distr:

$$p_k = 1/k, k \text{ outcomes }, 1 \leq k \leq K, H = log_2 K$$

so we get maximum entropy when uniform.

Example 23. Entropy of a geometric RV:

$$p_X(x) = p(1-p)^k$$

So: $H = -\sum pq^k log_2$

Entropy doesnt care about location: dice 1/2/3 values all havesame entropy as outcomes 4/5/6. Can get more interesting outcomes when probabilities are nonuniform. To calculate expected value, we can either directly calculate it, or calculate the entropy of an experiment to approximate distributions.
Before, people thought you could only increase signal by increasing power. Shannon showed that there was a different way; you can communicate over very noisy channels: call this "Shannon communication limited theory"

17 Thursday 4/10/2025

17.1 Random and Stochastic Processes

$X(t)$ is a RV as a function of time. each point has it's own pdf, cdf, Ψ, Φ, etc.

Example 24. $x(t) = Acos(\Omega t), A\ N(0,1), \Omega\ unif(0, 2\pi)$

$$Pr[X(t) \leq x] = F_{X(t)}(x; y), f_{X(t)}(x; t) = \frac{d}{dx} F_X(t)(x; t)$$

this is nothign new, just different notations.
1st order distributions: one point at a time:

$$F_{X(t)}(x; t) = Pr(X(t) \leq x), f_{X(t)}(x; t) = \frac{d}{dx} F_{X(t)}(x; t)$$

2nd order distributions: 2 points at a time:

$$X(t), X(\tau) \rightarrow \text{ can get 2D RV for a fixed } (t, \tau)$$

Similar for pdf and CDF, and kth order distributions. Note that we have to be careful with discrete values and discrete time.

In system theory, we introduce more constraints to understand the system. ie if input is $x(t)$, output $y(t)$, we can introduce linearity of the system to get the output is: $y(t) = \int h(t, \tau) x(\tau) d\tau = x(t) * h(t)$. LTI, then we can also use frequency response.
If we introduce independent increments, then we get a sequence of RVs. can do things usch as take a running subtotal of gaussian RVs: this is a **weiner process**.

Definition 21. $X(t)$ is a Markov process if $\forall k, t_1 < \cdots < t_k$, $P(X(t_k) = x_k | x(t_{k-1}) = Pr(x(t_k) = x_k | x(t_{k-1}, x(t_{k-2}), \ldots)$

In many important cases, we only need 1st or 2nd order statistics: mean and variance.

- **1st order**: average power on 1Ω basis: $\bar{P} = E(X^2(t))$; this is 2nd moment associated with power
- 2 points t, τ: have autocorrelation $R_X(t, \tau) = E(x(t)x(\tau))$ and autocovariance: $C_X(t, \tau) = E((x(t) - m_x(t))(x(\tau) - m_x(\tau))) = R_x(t, \tau) - m_x(t)m_x(\tau)$ a special case for 2 points is when $t = \tau$: get $R_x(t, t) = E(X^2(t))$; this is power of 1st order. autocovariance is the standard deviation.
- correlation covariance of 2 points:

$$\begin{aligned} \rho(t, \tau) &= \frac{C_X(t, \tau)}{\sqrt{C_x(t, t) C_x(\tau, \tau)}} \\ &= \frac{C_X(t, \tau)}{\sigma_X(t) \sigma_x \tau} \end{aligned}$$

- if we have iid, then autocorrelation is $\bar{A}^2$ if $n = m$, $\bar{A}^2$ else and autocovariance is 0 if not equal, else $var(A)$ for $n = m$.

Eventually, need to estimate statistics by taking samples, obtained through measurements.

Definition 22. *ensemble average*, simulating multiple times and taking the average of these simulations

Can do the same to get other statistics, also an ensemble histogram. Strong law of large numbers is an ergodic theorem.

Definition 23. An **Ergodic Theorem** states conditions under which time average converges as observation intervals become lrage

law of large numbers states as n grows large:

$$P[lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = m] = 1$$

We are estimating m by taking time average of realizations. If we look at the outer produce matrix: $x(t) \cdot x(\tau)$, averaging should give us an autocorrelation matrix for large N; this is "law of large numbers for autocorrelation". Important to note tehat in general, time averages do not converge to ensemble averages.

18 Tuesday 4/15/2025

Begin discussion on stochastic processes

Theorem 5 (Flip theorem)**.** Suppose we have a random variable $A \in \{-1, 1\}$ with equal probability. Let $X(t)$ be a stochastic process with mean $m(t)$ and autocorrelation $R_X(t, \tau)$. Define a new process $Y(t) = AX(t)$. Then:

$$E[Y(t)] = 0, \quad R_Y(t, \tau) = E[Y(t)Y^*(t + \tau)] = R_X(t, \tau)$$

In other words, the mean becomes zero, and the autocorrelation remains unchanged.

Now consider multiple random processes. Let $(X(t_1), \ldots, X(t_k))$ be independent of $(Y(\tau_1), \ldots, Y(\tau_k))$ for all sample locations t_i, τ_j. When dealing with two processes, we define the **cross-correlation**:

$$R_{XY}(t, \tau) = E[X(t)Y(\tau)]$$

In general, $Y(\tau)$ may be complex, so we consider:

$$R_{XY}(t, \tau) = E[X(t)Y^*(\tau)]$$

We also define the **cross-covariance**:

$$C_{XY}(t, \tau) = R_{XY}(t, \tau) - E[X(t)]E[Y^*(\tau)]$$

- **Orthogonal** if $R_{XY}(t, \tau) = 0$
- **Uncorrelated** if $C_{XY}(t, \tau) = 0$

Example 25. $x(t) = cos(\omega t + \theta), y(t) = sin(\omega t + \theta)$. Both mean of 0, what's cross correlation?

We can also model signals as random processes:

Example 26. signal + noise: $y(t) = x(t) + N(t)$
If X, N independent, what's R_{XN}?

Definition 24. $SNR = \sqrt{\frac{varX}{varN}}$: standard deviation ratios

For gaussian, uncorrelated is the same as independent. However, this is ONLY true for gaussian RVs
Looking at autocovariance of sum of processes:

$$S_n = \sum X[k] = x[n] + S_{n-1}$$

If $X[k]$ iid, $S_n = n\bar{X}$, $var(S_n) = n \cdot var(X)$.
For autocovariance:

$$\begin{aligned} C_S(n,k) &= E[(S_n - \bar{S_n})(s_k - \bar{s_k})] \\ &= E[(S_n - n\bar{X})(S_k - k\bar{X})] \\ &= E[(\sum(x_i - \bar{X}))(\sum(x_j - \bar{X}))] \end{aligned}$$

When $i = j$, solution is $var(X)$, otherwise zero; so how many such cases?

$$min(n,k) \to c_s(n,k) = min(n,k)var(X)$$

For the bernoulli case, $var(X) = pq, C_s(n,k) = min(n,k)pq$. Big n and p small, so $k << n$ since $p \approxeq \frac{k}{n} << 1$. therefore, n choose k is approximately $n^k/k!$ since n is large.
Now, $q^{n-k} = (1-p)^{n-k} \approxeq (1-p)^n \approxeq (e^{-p})^n$ (last step via taylor series approximation), so claim satisfied.
For a poisson process, we have the pdf

$$e^{-nt\tau} = \frac{(nt/\tau)^k}{k!}$$

if we let $n \to \infty$ s.t. $\lambda = n/T$: the freq of poitns remains constant, density of points is constant. λ is our RV average. Then:

$$Pr(kpointsonintervalt) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

For a weiner process, $var(X) = 4pq$ (bipolar case). if we expand the values/interval (ie from 0,1 to -1,1), variance grows as expected. So where can we use this? Poisson RV is one example: characterized by nubmer of occurrences over time period.

Definition 25. Poisson Approximation: for big K small p:

$$Pr(kpoints) = \binom{n}{k} p^k q^{n-k} \approxeq e^{-np}\frac{(np)^k}{k!}$$

the last approximateion is the claim: binomial to poisson to gaussian

19 Thursday 4/17/2025

Last time, the poisson process was defined by λ occurrences per unit time:

$$\text{Pr}[k \text{ points on interval } t] = P(K = k | T = t)$$

19.1 Poisson Counting Process

The occurrence of the events alone is not a stochastic process, but we can make it a stochastic process with a counting method. remember from poisson variable $\lambda = a$, mean/var are both a. Then, the counting process EV is:

$$E[X(t)] = \lambda t$$

If you plot the counting process, it will hover around the line λt.

The poisson counting process is an independent increment process. Therefore:

$$\begin{aligned}\forall \tau \geq t, j \geq i, Pr[x(t) = i, x(\tau) = j]& \\ &= Pr[x(t) = i, x(\tau) - x(t) = j - j] \\ &= Pr[x(t) = i] Pr[x(\tau) - x(t) = j - i] = \text{Pois}(t)\text{Pois}(t - \tau)\end{aligned}$$

Autocorrelation: if $\tau > t$,

$$\begin{aligned}R_x(t, \tau) &= E(x(t)x(\tau)) \\ &= E(x(t)(x(\tau) - x(t)) + x^2(t)) \\ &= E(x(t))E(x(\tau) - x(t)) + e(x^2(t)) = \lambda^2 t\tau + \lambda t\end{aligned}$$

To do a sanity check, simply check the units.
So what if $t > \tau$? just switch and calculate, we get:

$$\lambda^2 t\tau + \lambda \min(t, \tau)$$

For ambiguous t, τ.
First, let's look at autocovariance:

$$\begin{aligned}C_x(t, \tau) &= R_X() - E(x(t))E(x(\tau)) \\ &= \lambda^2 t\tau + \lambda \min(t, \tau) - \lambda t \lambda \tau \\ &= \lambda \min(t, \tau)\end{aligned}$$

Other RP related to the poisson process:**Random Telegraph Signal**

- Each poisson point flips from +1 to -1 (or alternative)
- $E = [x(t)] = e^{-2at}, C_x(t, \tau) = e^{-2a(t-\tau)}$
- we note for C_X, only thing that matters is distance between t, τ.

19.2 Poisson Point Process

the derivative of the poisson counting process results in dirac deltas:

$$Z(t) = \frac{d}{dt}x(t) = \sum \delta(t - s_n)$$

This is an important model for shot noise $v(t)$:

$$z(t) \rightarrow [h(t)] \rightarrow v(t) = \sum h(t - s_n)$$

Instead of δ, it's a signature function. Another sum process is with weiner process. In finance, when pricing securities S, we have: $dS = \mu S(t)dt + \sigma S(t)dV(t)$; the second part of the sum is considered the random portion, $V(t)$ a weiner process.

Definition 26 (Stationary Process)**. Stationary Process** similar to time invariance, does not change with respect to time

Have two types: strict stationary, and stationary in the wide sense (or **WSS**:wide-sense stationary)

Definition 27 (Strict Stationary)**. Strict stationary** if statistics of stochastic process are independent of where you choose your point: $F_{\vec{X}} = F^{\tau}_{\vec{X}}, \forall\tau$, ie any time shift τ, true for any number of samples

All i.i.d. processes are strictly stationary.

Definition 28 (Wide Sense Stationary)**. WSS** if the following are satisfied:

- mean constant for all time: $E(x(t)) = m$
- autocorrelation only a function of the distance between points: $R_x(t, \tau) = R_x(t - \tau)$

Some WSS properties: recall that $E(p(t)) = E(x^2(t)), R_X(t, \tau) = E(x(t)x(\tau))$, then:

$$R_x(t, t) = E(x^2(t)) = E(p(t))$$

so if $x(t)$ is WSS, then:

$$R_x(0) = E(x^2(t)) = E(p(t))$$

Assuming everything but θ is fixed in a stochastic process $x(t) = cos(\omega t + \theta)$, then:

$$\begin{aligned} &E(x(t)) = 0 \\ &R_x(t, \tau) = A^2/2cos(\omega(t - \tau)) = R_x(t - \tau) \end{aligned}$$

Then $x(t)$ is w.s.s. with $R_x(t) = A^2/2cos(\omega t)$, and we have both conditions satisfied, with $R_x(0) = E(p(t)) = A^2/2$. Recall voltage of sinusoid is $A/\sqrt{2}$. This can be estimated emperically, as was done in the past.
Some more properties:

- $R_x(0) = E(x^2(t))$
- $R_x(\tau) = R_x(-\tau)$
- $R_x(\tau) \leq R_x(0)$
- if $R_x(\tau) = R_x(0)$ for some $\tau > 0$, then $R_x(\tau)$ is periodic

Note even function in second bullet point; same if complex. third bullet point makes sense because a number is most correlated with itself.
In conclusion, we discussed stationary stochastic processes, in both strict and wide sense. the WSS is similar to an LTI. We will eventually take W.S. Stochastic process and run it through and LTI with freq response $h(t)$: $x(t) \to [h(t)] \to Y(t)$; if $x(t)$ is WSS then so is $Y(t)$. also, $R_x(t) \leftrightarrow S_X(\omega)$, the power spectrum density (PSD).

20 Tuesday 4/22/2025

multidimentional gaussian RV:

$$\text{PDF: } f_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{k/2}|K|^{1/2}}$$

K is correlation matrix.

$$\text{Characteristic Function: } \Phi_{\vec{X}}(x) = E(\ldots.) = e^{j\vec{\omega}\frac{1}{2}\vec{\omega}^T K\vec{\omega}}$$

Theorem 6. If Gaussian RP is WSS (wide-sense stationary), then it's also SSS (strict-sense stationary).

Proof. WSS implies $E(x(t)) = m$ for all t by definition, also $C_x(t, \tau) = C_x(t - \tau)$.
Choose k points: $\{t_i | 1 \leq i \leq k\}$. then the elements of K matrix are:

$$(K)_{nm} = C_X(t_n, t_m) = C_x(t_n - t_m)$$

Thus, the char. function is:

$$\begin{aligned}\Phi_X(x) &= exp(j\sum_i = 1^k \omega_i E(x_i) - \frac{1}{2}\sum_1^k sum_1^k \omega_n \omega_m (K)_{nm} \\ &= exp(jm\sum \omega_i - \frac{1}{2}\sum\sum \omega_n \omega_m C_X(t_n - t_m)) \text{ (by substitution)}\end{aligned}$$

Now, take the k points, add τ, and verify it's the same PDF:

$$\Phi_{X(t-\tau)}(\vec{\omega}) = exp(\ldots C_X((t_m - \tau) - (t_n - \tau)) = \Phi_{X(\vec{t})}(\vec{\omega})$$

$\square$

Usually, strict sense is a subset of wide-sense. Theyre the same for gaussian RVs.

20.1 Ergodic Theorems

briefly mentioned prior.

- Computing parameters of stochastic process
- before, we took means of ensembles: $\frac{1}{k}\sum_1^k X_i(t) \to \bar{X(t)}$
- however, many times, we don't have luxury of ensemble, we only have 1 long realization. Intuitively, we're limited to **stationary processes**

Definition 29. Mean Ergodic: The time-averaged value of a process $X(t)$ over an interval of length $2T$ is defined as

$$\langle X(t)\rangle_T = \frac{1}{2T}\int_{-T}^{T} X(t)\,dt$$

As $T \to \infty$, the time average converges to the ensemble average:

$$\langle X(t)\rangle_T \to \bar{X}$$

We may be able to get mean but not other statistics guaranteed. If we can get the variance, we call it *variance ergodic*

Definition 30. If a parameter can be found from a single observation of the process, the process is called **ergodic**

Another idea is that we can break down $X(t)$ into multiple processes and uses them to form an ensemble

Example 27. not all processes are mean ergodic: take a battery factory. Each sample batter has it's own average:

$$X_1(t) = 1.48$$
$$X_2(t) = 1.52$$
$$X_3(t) = 1.51$$

Can assume stochastic process is constant in time, add all up and average for an average battery voltage.
Can you take a single realization to get the mean??? no! Each realization is only the fixed value. Therefore, battery factory is not mean ergodic.

Just because we can perform and ensemble mean doesnt mean a process is a mean ergodic process. Let's assume $X(t)$ is wss:

$$E[\langle x(t)\rangle_T] = \frac{1}{2T}\int_{-T}^{T} \bar{X}dt = \bar{X}$$

We would like variance to approach zero as $t \to \infty$. If so, then $x(t)$ mean ergodic:

$$var(\dots) = E[(\langle x(t)\rangle_T - \bar{x})^2] = E[(\frac{1}{2}\int_{-T}^{T}(x(t)-\bar{x})dt)(\frac{1}{2T}\int_{-T}^{T}(x(\tau)-\bar{x})d\tau)]$$
$$= \frac{1}{4T^2}\int\int C_X(t-\tau)dtd\tau = \frac{1}{4\pi^2}\int\int C_X(t-\tau)\Pi(\frac{\tau}{2T})\Pi(\frac{t}{2T})dtd\tau$$

using $\xi = t - \tau$ substitute, and we get defintion of convolution between $(\Pi * \Pi) \to \Lambda$. Thus we have:

$$var(\dots) = \frac{1}{2T}\int_{\mathbb{R}} C_X(\tau)\Lambda(\frac{\tau}{2T})d\tau$$

And so we get variance becomes 0 as $\tau \to 0$. This leads to lemma:

Lemma 1. A WSS RP is mean ergodic if:

$$\lim_{\tau\to\infty}\frac{1}{2T}\int_{\mathbb{R}} C_X(\tau)\Lambda(\frac{\tau}{2T})d\tau = 0$$

Example 28. Define telegraph signal: bunch of poison points: alternate between +1, -1; have:

$$C_X(\tau) = e^{-2\lambda(\tau)}$$

So further from point, the smalle autocovariance becomes.
Is this mean ergodic? yes; can show by looking at criteria

Example 29. battery factory, know there's a mean and variance.

$$\to C_X(x) = var(x) \to \frac{1}{2\pi}\int C_X(\tau)\Lambda d\tau = var(x)\frac{1}{\tau}\int(1-\frac{\tau}{2T})d\tau$$

so as $\tau \to \infty$, var does not go to 0

20.2 Autocovariacne Ergodic

For large T, can we approximate

$$R_X(\tau) = \langle x(t)x(t-\tau)\rangle_T$$

assume WSS: for fixed τ, define new stochastic process:

$$Y_\tau(t) = X(t)X(t-\tau)$$

If $Y(t)$ is mean ergodic, then product is also ergodic. reducing to 1D for a fixed τ. Note:

$$C_{Y_\tau}(\xi) = E[y_\tau(t)Y_\tau(t-\xi)] = E[x(\alpha)x(\alpha+t)x(\beta)x(\beta+t+\xi)]$$

Here, we need to look at 4th order statistics. the point is that ergodicity can get messy: this is more than just WSS.

21 Thursday 4/24/2025

Autocorrelation of a WSS process as the following FT:

$$R_X(\tau) \leftrightarrow S_X(f) : \text{ this is the } \textbf{power spectrum density}$$

for an LTI: we know

$$X(t) \to [h(t)] \to Y(t) = X(t) * h(t)$$

$Y(t)$ is also a WSS and stationary if $X(t)$ is a WSS. Turns out:

$$\text{(PSD of output) } S_Y(t) = S_X(f)|H(f)|^2$$

$H(f)$ the frequency response of the system. How about an input through a diode? Diode is not an LTI but it is TI, so we'd expect yes, WSS is still a stochastic process.
But what is the PDF? PDF, integrate and you get probability, Mass DF you get mass, so we'd expect if you integrate, you get the power. also related to frequency due to "...spectrum..." term. As stated, PSD of WSS process is the fourier transform of its autocorrelation:

$$S_X(f) = \int_\tau R_X(\tau)e^{-j2\pi f\tau}d\tau$$

By FT property $\int x(t) = X(0)$, then:

$$\int S_X(f)df = R_X(0) = P = E(|x(t)|^2)$$

So if we have PSD, we can integrate to get average power. Power in specified freqeuncy band can be obtained by obtaining the PSD:

$$\text{Power in band } (B_L < f < B_U) = 2\int_{f=B_L}^{B_U} S_X(f)df$$

note, the 2 appears because the FT is complex, so have to multiple by 2 to account for the negative frequencies

Example 30. random telephon signal:

$$R_X = e^{-2a|\tau|} \leftrightarrow S_X(t) = \frac{a}{a^2 + (\pi f)^2}$$

Entire power is $\int S_X = R_X(0) = 1$, and power in band is:

$$= 2\int_{B_L}^{B_U} S_X = \cdots = \frac{2}{\pi}(tan^{-1}(\frac{\pi B_U}{a}) - tan^{-1}(\frac{\pi B_L}{a}))$$

We like taking BP filter of process, we get the above power in that band.

PSD of discrete R.P. Let $x(n)$ be WSS. then the PSD is DTFT of autocorrelation:

$$S_X(n) = \sum_n R_X[n]e^{j2\pi nu}$$

The term is the fourier series with a period of 1. So we only need to look at 1 period of DTFT.

21.1 Types of Noise

- continuous white noise:
$$R_X(\tau) = \frac{N_0}{2}\delta(\tau), S_X(u) = \frac{N_0}{2}$$
  Problem is that $\int S_X = \infty$
- thermal noise similar ot white noise
- white gaussian noise
  "white" refers to the spectral density
  "gaussian" refers to what happens if you sample the noise
- discrete white noise
  this actually does exist:
$$R_X[n] = \bar{x^2}\delta][n]$$
$$S_X(f) = \bar{X}^2$$

an example is i.i.d. of sequences of random variables

22 Tuesday 4/29/2025

We have for a system with respond $H(t)$:

$$S_X(t) \to [H(f)] \to S_Y(f) = S_X(f)|H(f)|^2$$

for $Y(t) = \frac{d}{dt}X(t)$, $H(f) = j2\pi f$ due to the derivative property.
A refersher on linear systems and LSIs:

- homogeniety: if $y = sx$, then $cSx = Scx$
- additivity: $S(x_1) + S(x_2) = S(x_1 + x_2)$

FT is linear but has no frequency response.
The derivative of the *superposition integral*:
Let $x(t) = \int_{-\infty}^{infty} x(\tau)\delta(\tau - t)d\tau$, then

$$Sx(t) = S\int x(\tau)\delta(\tau - t)d\tau$$

by additivity of the integral:

$$= \int S(x(\tau)\delta(\tau - t))d\tau = \int x(\tau)S(\delta(\tau - t))d\tau = h(t;\tau)$$

Shifted deltas may not have the same output. note:

$$y(t) = \int_{-\infty}^{i} nftyx(\tau)e^{-j2\pi t\tau}d\tau$$

if FT is a system, then the impulse response of the FT is:

$$h(t,\tau) = e^{-j2\pi t\tau}$$

but we need an LTI and not just linearity (ie diode is not lti but it is TI)
time invariant:
If $y(t) = Sx(t)$, then $Sx(t-\tau) = y(t-\tau)$. Properties: linearity means homogeneity and additivity, time invariant means if $y(t) = Sx(t)$, then $y(t-\tau) = Sx(t-\tau)$; LTI is the intersection of these two. A third system: *causal system*. All temporal sytems are causal, no spatial systems are causal. Note the similarity between time-invariance and stationarity; $R(t;\tau) = R(t-\tau)$. So what happens if we put stochastic process through a linear system, or even a linear time system?

22.1 Through linear system

what happens to expectations, correlations?

$$E(y(t)) = E(L(x(t))) = L(E(x(t)))$$

If LTI, then $y(t) = h(t) * x(t)$.
Let's look at crosscorrelation between input adn output:

$$\begin{aligned}
R_{XY}(t,\tau) &= E(x(t)y(\tau)) \\
&= E(x(t)(x(\tau) * h(\tau)) \\
&= E(x(t) \int x(\tau)h(t-\tau)d\tau \\
&= R_X(t,\tau) * h(\tau)
\end{aligned}$$

How about $R_Y(t,\tau)$?

$$\begin{aligned}
&= E(y(t)y(\tau)) \\
&= E((x(t) * h(t))y(\tau)) \\
&= E(\int x(a)h(t-a)day(\tau)) \\
&h(t) * R_{XY}(t,\tau)
\end{aligned}$$

This says we're going to cross convolve with $h(t)$ for each row (the stochastic process).
note: LTI but the random process may not be stationary:

$$\begin{aligned}
R_Y &= h(t) * R_{XY}(t) \\
R_{XY} &= R_X * h(t) \\
R_Y &= h(t) * R_{XY}(t,\tau) = h(t) * R\ldots
\end{aligned}$$

How about different random processes? $\frac{d}{dt}$ also lower, but no defined impulse response:

$$h(t) = S\delta(t) = \frac{d}{dt}\delta(t)$$

Don't deal with this, but can deal with H(f):

$$\begin{aligned}
y(t) &= \frac{dx}{dt}(t) \\
R_{XY} &= \frac{d}{d\tau}R_X \\
R_y &= \frac{d}{dt}R_{XY} \\
R_y &= \frac{d^2}{dtd\tau}R_X
\end{aligned}$$

This leads to stochastic differential equations (ie Maxwell's equation)

$$\sum b_n \frac{b^n}{dt^n}y(t) = \sum a_n \frac{d^m}{dt^m}x(t)$$

What happens if input is WSS? Mean is also WSS:

$$\begin{aligned}
E(y(t)) &= E(\int x \ldots h) \\
&= \int E(x)h(\ldots)d\tau \\
&= m_x \int h(\tau)d\tau \\
&= m_x H(0)
\end{aligned}$$

Output $y(t)$ is also WSS; can show the autocorrelation is not a function of t, and hence WSS. Lastly, properties of LSI, WSS input:

$$\begin{aligned}
&S_Y(F) = S_X(f)|H(f)|^2 \\
&E(y(t)) = m_x H(0)
\end{aligned}$$

23 Thursday 5/1/2025

Sampling theorem: $x(t) \leftrightarrow X(u)$, $x(t)$ band-limited, so can sample at $\frac{1}{2B}$:

$$x(t) = \sum_n x(\frac{n}{2B})sinc(2Bt - n), sinc(t) = \frac{sin(\pi t)}{\pi t}$$

each sample has a corresponding sinc: 0 at every other point. So what happens if you add noise to the sample? $y(t) = x(t) + \xi(t)$:

$$\begin{aligned}
z(t) &= \sum_n y(\frac{n}{2B})sinc(2Bt - n) \\
&= \sum_n (x(\frac{n}{2B}) + \xi(\frac{n}{2B})sinc(2Bt - n)) \\
&= x(t) + \nu(t), \nu(t) = \sum_n \xi(\frac{n}{2B})sinc(2Bt - n)
\end{aligned}$$

Let's look at the autocorrelation:

$$\nu(t)\nu(\tau) = \sum_q \nu(\frac{q}{2B})sinc \cdots \sum_q \cdots = \sum_q \sum_p$$

take the expectation:

$$R_\nu(t,\tau) = \nu(t)\bar{\nu}(\tau) = \sum_p \sum_p \xi(\frac{p}{2B})\bar{\xi}(\frac{q}{2B})$$

If we assume ξ is WSS, then:

$$\rightarrow \sum_p \sum_q R_\xi(\frac{p-w}{2B})sinc(2Bt-q)sinc(2B\tau-p)$$

let $n = p - q \rightarrow q = p - n$ by substitution:

$$\rightarrow \sum_n \sum_p R_\xi(\frac{n}{2B})sinc(2Bt-p+u)sinc(2B\tau-p) = \sum_n R_\xi(\frac{n}{2B}) \sum_p sinc(\dots)sinc(\dots)$$

note that sum_p part is independent of the noise, and $since(2Bt-p)$ is kernel of the sampling theorem. let's interpret $since(2Bt-p+u)$ as the function we sample. since $sinc(2B(t-\tau)+n) = sinc(2B(\tau-t)-\xi)$ is bandlimited, expanding the sampling theorem:

$$\rightarrow = \sum_p sinc(2Bt-p+n)sinc(2B\tau-p)$$

$$\rightarrow R_Y(t,\tau) = \sum R_\xi()sinc() = R_\nu(\tau-t) = R_\nu(t-\tau)$$
$$\rightarrow R_\nu(\tau) = \sum_n R_\xi(\frac{n}{2B})sinc(2B\tau-n)$$

Back to the power spectral density:

$$R_\nu(\tau) = \sum R_\xi(\frac{n}{2B})sinc(2B\tau-n)$$
$$\leftrightarrow$$
$$S_n(f) = \frac{1}{2B}(\sum_n R_\xi(\frac{n}{2B})e^{-j2\pi\frac{n}{2B}})\Pi(\frac{f}{2B})$$

Remember the sampling theory is the fourier dual of the fourier series.
What if noise sample is white noise (ie uncorrelated)? then:

$$R_\xi(\frac{n}{2B}) = \bar{\xi^2}\delta[n]$$

then:

$$R_\nu(\tau) = \bar{\xi^2}sinc(2B\tau)$$
$$\rightarrow @_\nu(f) = \frac{\bar{\xi^2}}{2B}\Pi(\frac{f}{2B}) : \text{ BL white noise}$$