Warning: This lecture is rated R.

1. Conditional Expectation
   - Review
   - Going Viral
   - Walt’s Identity
   - CE = MMSE

2. Continuous Probability
   - Motivation.
   - Continuous Random Variables.
   - Cumulative Distribution Function.
   - Probability Density Function
   - Expectation and Variance
**Definition** Let $X$ and $Y$ be RVs on $\Omega$. The conditional expectation of $Y$ given $X$ is defined as

$$E[Y|X] = g(X)$$

where

$$g(x) := E[Y|X = x] := \sum_y yPr[Y = y|X = x].$$
Properties of Conditional Expectation

\[ E[Y|X = x] = \sum_y y Pr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends. Each of your friend retweets w.p. $p$.
Each of your friends has $d$ friends, etc.
Does the rumor spread? Does it die out (mercifully)?

In this example, $d = 4$. 
**Application: Going Viral**

**Fact:** Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

**Proof:**

Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1} | X_n = k] = kpd$.

Thus, $E[X_{n+1} | X_n] = pdX_n$. Consequently, $E[X_n] = (pd)^{n-1}, n \geq 1$.

If $pd < 1$, then $E[X_1 + \cdots + X_n] \leq (1 - pd)^{-1} \implies E[X] \leq (1 - pd)^{-1}$.

If $pd \geq 1$, then for all $C$ one can find $n$ s.t. $E[X] \geq E[X_1 + \cdots + X_n] \geq C$.

In fact, one can show that $pd \geq 1 \implies Pr[X = \infty] > 0$. 
Application: Going Viral

An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1}|X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k).$$

Thus, $E[X_{n+1}|X_n = k, D_1, \ldots, D_k] = p(D_1 + \cdots + D_k)$.

Consequently, $E[X_{n+1}|X_n = k] = E[p(D_1 + \cdots + D_k)] = pdk$.

Finally, $E[X_{n+1}|X_n] = pdX_n$, and $E[X_{n+1}] = pdE[X_n]$.

We conclude as before.
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where $Z$ takes values in $\{0, 1, 2, \ldots\}$ and $E[X_n] = \mu$ for all $n \geq 1$.

Then,

$$E[X_1 + \cdots + X_Z] = \mu E[Z].$$

**Proof:**

$E[X_1 + \cdots + X_Z | Z = k] = \mu k$.

Thus, $E[X_1 + \cdots + X_Z | Z] = \mu Z$.

Hence, $E[X_1 + \cdots + X_Z] = E[\mu Z] = \mu E[Z]$. 

$\square$
CE = MMSE

**Theorem**

$E[Y|X]$ is the ‘best’ guess about $Y$ based on $X$.

Specifically, it is the function $g(X)$ of $X$ that minimizes $E[(Y - g(X))^2]$. 
**Theorem** $CE = MMSE$

$g(X) := E[Y|X]$ is the function of $X$ that minimizes

$$E[(Y - g(X))^2]$$

**Proof:**

First recall the projection property of CE:

$$E[(Y - E[Y|X])h(X)] = 0, \forall h(\cdot).$$

That is, the error $Y - E[Y|X]$ is *orthogonal* to any $h(X)$. 
CE = MMSE

**Theorem** CE = MMSE

\( g(X) := E[Y|X] \) is the function of \( X \) that minimizes

\[ E[(Y - g(X))^2] \]

.  

**Proof:** Let \( h(X) \) be any function of \( X \). Then

\[
E[(Y - h(X))^2] = E[(Y - g(X) + g(X) - h(X))^2] \\
= E[(Y - g(X))^2] + E[(g(X) - h(X))^2] \\
+ 2E[(Y - g(X))(g(X) - h(X))].
\]

But,

\[
E[(Y - g(X))(g(X) - h(X))] = 0 \text{ by the projection property.}
\]

Thus, \( E[(Y - h(X))^2] \geq E[(Y - g(X))^2] \).
\(E[Y|X]\) and \(L[Y|X]\) as projections

\(L[Y|X]\) is the projection of \(Y\) on \(\{a + bX, a, b \in \mathbb{R}\}\): LLSE

\(E[Y|X]\) is the projection of \(Y\) on \(\{g(X), g(\cdot): \mathbb{R} \rightarrow \mathbb{R}\}\): MMSE.
Escapes from SPECTRE sometime during 1,000 mile flight.
Uniformly likely to be at any point along path.

What is the chance he is at any point along the path?
Discrete Setting: Uniform over $\Omega = \{1, \ldots, 1000\}$.
Continuous setting: probability at any point in $[0, 1000]$? Probability at any one of an infinite number of points is .. ...uh ...0?
Consider $[a, b] \subseteq [0, \ell]$ (for James, $\ell = 1000$.)
Let $[a, b]$ also denote the event that point is in the interval $[a, b]$.

$$Pr[[a, b]] = \frac{\text{length of } [a, b]}{\text{length of } [0, \ell]} = \frac{b - a}{\ell} = \frac{b - a}{1000}.$$ 

Again, $[a, b] \subseteq \Omega = [0, \ell]$ are events.
Events in this space are unions of intervals.
Example: the event $A$ - “within 50 miles of base” is $[0, 50] \cup [950, 1000]$.

$$Pr[A] = Pr[[0, 50]] + Pr[[950, 10000]] = \frac{1}{10}.$$
Another Bond example: Spectre is chasing him in a buggie. Bond shoots at buggy and hits it at random spot. What is the chance he hits gas tank? Gas tank is a one foot circle and the buggy is 4 × 5 rectangle.

$$\Omega = \{(x, y) : x \in [0, 4], y \in [0, 5]\}.$$  

The size of the event is $$\pi(1)^2 = \pi.$$  

The “size” of the sample space which is 4 × 5. Since uniform, probability of event is $$\frac{\pi}{20}.$$
Buffon’s needle.

Throw a needle on a board with horizontal lines at random.

Lines 1 unit apart, needle has length 1. What is the probability that the needle hits a line? Clearly...

\[
\frac{2}{\pi}.
\]
Buffon’s needle.

Sample space: possible positions of needle.
Position: center position \((X, Y)\), orientation, \(\Theta\).

Relevant: \(X\) coordinate doesn’t matter; \(Y\) coordinate := distance from closest line. \(Y \in [0, \frac{1}{2}]\); \(\Theta :=\) closest angle to vertical \([-\frac{\pi}{2}, \frac{\pi}{2}]\). When \(Y \leq \frac{1}{2} \cos \Theta\): needle intersects line.

\[
Pr[\text{“intersects”}] = \int_{-\pi/2}^{\pi/2} \left( Pr[\Theta \in [\theta, \theta + d\theta]] Pr[Y \leq \frac{1}{2} \cos \theta] \right) d\theta \\
= \int_{-\pi/2}^{\pi/2} \left( \frac{d\theta}{\pi} \times \left[ \frac{1}{2} \cos \theta \right] \right) = \frac{2}{\pi} \left[ \frac{1}{2} \sin \theta \right]_{-\pi/2}^{\pi/2} = \frac{2}{\pi}.
\]
Continuous Random Variables: CDF

Pr[a ≤ X ≤ b] instead of Pr[X = a].
For all a and b specifies the behavior!
Simpler: P[X ≤ x] for all x.

Cumulative probability Distribution Function of X is

\[ F(x) = Pr[X ≤ x] \]

Pr[a < X ≤ b] = Pr[X ≤ b] − Pr[X ≤ a] = F(b) − F(a).
Idea: two events X ≤ b and X ≤ a.
Difference is the event a ≤ X ≤ b.
Example: CDF

Example: Bond’s position.

\[ F(x) = Pr[X \leq x] = \begin{cases} 
0 & \text{for } x < 0 \\
\frac{x}{1000} & \text{for } 0 \leq x \leq 1000 \\
1 & \text{for } x > 1000 
\end{cases} \]

Probability that Bond is within 50 miles of center:

\[ Pr[450 < X \leq 550] = Pr[X \leq 550] - Pr[X \leq 450] \]

\[ = \frac{550}{1000} - \frac{450}{1000} \]

\[ = \frac{100}{1000} = \frac{1}{10} \]
Example: CDF

Example: hitting random location on gas tank.
Random location on circle.

Random Variable: $Y$ distance from center.
Probability within $y$ of center:

$$Pr[Y \leq y] = \frac{\text{area of small circle}}{\text{area of dartboard}}$$

$$= \frac{\pi y^2}{\pi} = y^2.$$

Hence,

$$F_Y(y) = Pr[Y \leq y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$
Calculation of event with dartboard.

Probability between .5 and .6 of center?
Recall CDF.

\[ F_Y(y) = \Pr[Y \leq y] = \begin{cases} 
  0 & \text{for } y < 0 \\
  y^2 & \text{for } 0 \leq y \leq 1 \\
  1 & \text{for } y > 1 
\end{cases} \]

\[
\Pr[0.5 < Y \leq 0.6] \quad = \quad \Pr[Y \leq 0.6] - \Pr[Y \leq 0.5] \\
\quad = \quad F_Y(0.6) - F_Y(0.5) \\
\quad = \quad .36 - .25 \\
\quad = \quad .11
\]
Density function.

Is the dart more like to be (near) .5 or .1?
Probability of “Near x” is $Pr[x < X \leq x + \delta]$.
Goes to 0 as $\delta$ goes to zero.
Try

$$\frac{Pr[x < X \leq x + \delta]}{\delta}.$$  

The limit as $\delta$ goes to zero.

$$\lim_{\delta \to 0} \frac{Pr[x < X \leq x + \delta]}{\delta} = \lim_{\delta \to 0} \frac{Pr[X \leq x + \delta] - Pr[X \leq x]}{\delta}$$

$$= \lim_{\delta \to 0} \frac{F_X(x + \delta) - F_X(x)}{\delta}$$

$$= \frac{d(F(x))}{dx}.$$
Definition: (Density) A probability density function for a random variable $X$ with cdf $F_X(x) = Pr[X \leq x]$ is the function $f_X(x)$ where

$$F_X(x) = \int_{-\infty}^{x} f_X(x) \, dx.$$ 

Thus,

$$Pr[X \in (x, x + \delta)] = F_X(x + \delta) - F_X(x) = f_X(x) \delta.$$
Examples: Density.

Example: uniform over interval $[0, 1000]$

$$ f_X(x) = F_X'(x) = \begin{cases} 
0 & \text{for } x < 0 \\
\frac{1}{1000} & \text{for } 0 \leq x \leq 1000 \\
0 & \text{for } x > 1000 
\end{cases} $$

Example: uniform over interval $[0, \ell]$

$$ f_X(x) = F_X'(x) = \begin{cases} 
0 & \text{for } x < 0 \\
\frac{1}{\ell} & \text{for } 0 \leq x \leq \ell \\
0 & \text{for } x > \ell 
\end{cases} $$
Examples: Density.

Example: “Dart” board.
Recall that

\[
F_Y(y) = Pr[Y \leq y] = \begin{cases} 
0 & \text{for } y < 0 \\
y^2 & \text{for } 0 \leq y \leq 1 \\
1 & \text{for } y > 1 
\end{cases}
\]

\[
f_Y(y) = F'_Y(y) = \begin{cases} 
0 & \text{for } y < 0 \\
2y & \text{for } 0 \leq y \leq 1 \\
0 & \text{for } y > 1 
\end{cases}
\]

The cumulative distribution function (cdf) and probability distribution function (pdf) give full information.
Use whichever is convenient.
$U[a, b]$
The exponential distribution with parameter $\lambda > 0$ is defined by

$$f_X(x) = \lambda e^{-\lambda x} 1\{x \geq 0\}$$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases}$$
Recall that $Pr[X \in (i\delta, i(\delta + 1))] = f_X(i\delta)\delta$. Thus,

$$E[X] = \sum_{i=-\infty}^{\infty} (i\delta)Pr[i\delta < X \leq (i + 1)\delta]$$

$$= \sum_{i=-\infty}^{\infty} (i\delta)f_X(i\delta)\delta$$

$$= \int_{-\infty}^{\infty} xf_X(x)dx.$$
Expectation of $U[a, b]$

Let $X = U[a, b]$. That is,

\[ f_X(x) = \frac{1}{b - a} 1 \{ a \leq x \leq b \}. \]

Hence,

\[
E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx = \int_{a}^{b} x \frac{1}{b - a} \, dx
\]

\[
= \frac{1}{b - a} \int_{a}^{b} x \, dx = \frac{1}{b - a} \left[ \frac{x^2}{2} \right]_{a}^{b}
\]

\[
= \frac{1}{2(b - a)} [b^2 - a^2] = \frac{a + b}{2}.
\]
Expectation: dartboard.

Example: distance from center on radius 1 dartboard.
Recall:

\[ f_Y(y) = 2y1\{0 \leq y \leq 1\}. \]

Hence,

\[
\int_{-\infty}^{\infty} y f(y) \, dy = \int_{-\infty}^{0} 0 + \int_{0}^{1} 2y^2 \, dy + \int_{1}^{\infty} 0 \, dy
\]

\[
= 0 + \frac{2y^3}{3}\bigg|_{0}^{1} + 0
\]

\[
= \frac{2}{3}
\]

Try whole process for general radius. What do you get?
Expectation: Exponential.

Let $X = \text{Expo}(\lambda)$

Then,

$$E[X] = \int_{-\infty}^{\infty} xf_X(x) \, dx = \int_{0}^{\infty} x\lambda e^{-\lambda x} \, dx$$

$$= -\int_{0}^{\infty} xde^{-\lambda x}$$

$$= (*) - \{[xe^{-\lambda x}]_{0}^{\infty} - e^{-\lambda x} \, dx\}$$

$$= \int_{0}^{\infty} e^{-\lambda x} \, dx = -\frac{1}{\lambda} \int_{0}^{\infty} de^{-\lambda x}$$

$$= -\frac{1}{\lambda} [e^{-\lambda x}]_{0}^{\infty} = \frac{1}{\lambda}.$$

(*) We used the integration by parts formula:

$$\int_{a}^{b} f(x)dg(x) = [f(x)g(x)]_{a}^{b} - \int_{a}^{b} g(x)df(x),$$

which follows from $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$.
Variance

**Definition:** The variance of a continuous random variable $X$ is

$$E((X - E(X))^2) = E(X^2) - (E(X))^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \left( \int_{-\infty}^{\infty} x f(x) dx \right)^2.$$ 

Example: uniform on $[0, \ell]$.

$$\int_{0}^{\ell} x^2 \frac{1}{\ell} dx = \frac{x^3}{3\ell} \bigg|_{0}^{\ell} = \frac{\ell^2}{3}.$$ 

And, $E(X) = \frac{\ell}{2}$. So

$$Var(X) = \frac{\ell^2}{3} - \frac{\ell^2}{4} = \frac{\ell^2}{12}.$$ 

$\approx \frac{n^2 - 1}{12}$ for uniform discrete distribution on $\{1, \ldots, n\}$.
Summary

1. \( E[Y|X] := \sum_y yPr[Y = y|X = x] \).
2. Properties: Linearity, ...., MMSE.
3. Applications: Diluting, Mixing, Going Viral, Wald.
4. Motivation for Continuous Probability: The world is continuous ....
5. pdf: \( Pr[X \in (x, x+\delta)] = f_X(x)\delta \).
6. CDF: \( Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy \).
7. \( U[a, b], \text{Expo}(\lambda) \), target.
8. Expectation: \( E[X] = \int_{-\infty}^{\infty} xf_X(x)dx \).