Conditional Expectation

**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

$$g(x) := E[Y|X = x] := \sum_y y \Pr[Y = y|X = x].$$

**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) $E[aY + b|X] = aE[Y|X] + bE[Z|X]$;
(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, $d = 4$.

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

Proof:

Given $X_0 = k, X_{n+1} = B(kd, p)$. Hence, $E[X_n|X_0 = k] = kpd$.

Thus, $E[X_n+1|X_0] = pdX_n$. Consequently, $E[X_n] = (pd)^n \cdot n \geq 1$.

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

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 $pd \geq 1 \Rightarrow \Pr[X = \infty] > 0$.

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, $E[X_1 + \cdots + X_2] = \mu E[Z]$.

**Proof:**
$E[X_1 + \cdots + X_2] = \mu k$.
Thus, $E[X_1 + \cdots + X_2|Z = k] = \mu k$.
Hence, $E[X_1 + \cdots + X_2] = E[\mu k] = \mu E[Z]$.

---

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

**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 $h(X)$.

But,
$E[(Y - g(X))(g(X) - h(X))] = 0$ by the projection property.
Thus, $E[(Y - h(X))^2] \geq E[(Y - g(X))^2]$.

---

**CE = MMSE**

$E[Y|X]$ 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 $\{g(X), g(\cdot) : \mathbb{R} \to \mathbb{R}\}$: MMSE.

---

**Continuous Probability - James Bond.**

- 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 .....

...
Continuous Probability: the interval!

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 \([0, 50] \cup [950, 1000]\).

\[
Pr[A] = Pr[[0, 50]] + Pr[[950, 1000]] = \frac{1}{10}
\]

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 chance he hits gas tank?

\[
\text{Gas tank is a one foot circle and the buggy is } 4 \times 5 \text{ rectangle.}
\]

Shooting..

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 \(4 \times 5\) rectangle.

---

Continuous Random Variables: CDF

\(Pr[a \leq X \leq b]\) instead of \(Pr[X = a]\).

For all \(a\) and \(b\) specifies the behavior!
Simpler: \(P[X \leq x]\) for all \(x\).

Cumulative probability Distribution Function of \(X\) is

\[
F(x) = Pr[X \leq x]
\]

\[
Pr[a < X \leq b] = Pr[X \leq b] - Pr[X \leq a] = F(b) - F(a).
\]

Idea: two events \(X \leq b\) and \(X \leq a\).
Difference is the event \(a < X \leq b\).

---

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:

\[
Pr[450 < X \leq 550] = Pr[X \leq 550] - Pr[X \leq 450] = \frac{550}{1000} - \frac{450}{1000} = \frac{100}{1000} = \frac{1}{10}
\]
**Density function.**

Is the dart more like to be (near) .5 or .1?
Probability of “Near x” is $Pr[x < X ≤ x + δ]$.

Try

$$Pr[x < X < x + δ] = \lim_{δ \to 0} \frac{Pr[X ≤ x + δ] − Pr[X ≤ x]}{δ} = \lim_{δ \to 0} \frac{F_X(x + δ) − F_X(x)}{δ} = d\left(\frac{dF(x)}{dx}\right)$$

The limit as $δ$ goes to zero.

**Examples: Density.**

Example: “Dart” board.
Recall that

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

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

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

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

$$f_X(x) = F_X'(x) = \begin{cases} 
0 & \text{for } x < 0 \\
y & \text{for } 0 ≤ x ≤ ℓ \\
0 & \text{for } x > ℓ 
\end{cases}$$

The cumulative distribution function (cdf) and probability distribution function (pdf) give full information.

Use whichever is convenient.

**Calculation of event with dartboard.**

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

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

$$Pr[0.5 < Y ≤ 0.6] = Pr[Y ≤ 0.6] − Pr[Y ≤ 0.5] = F_Y(0.6) − F_Y(0.5) = .36 − .25 = .11$$

**Definition: (Density) A probability density function for a random variable $X$ with cdf $F_X(x) = Pr[X ≤ x]$ is the function $f_X(x)$ where

$$F_X(x) = \int_{-∞}^{x} f_X(x)dx.$$**

Thus,

$$Pr[X ∈ (x, x + δ)] = F_X(x + δ) − F_X(x) = f_X(x)δ.$$
**Uniform in \([a, b]\)**

Let \(X = U[a, b]\). That is, 
\[
f_X(x) = \frac{1}{b-a} 1\{a \leq x \leq b\}.
\]
Hence, 
\[
E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b x \cdot \frac{1}{b-a} \,dx = \frac{1}{b-a} \int_a^b x \,dx = \frac{1}{2(b-a)} \left[ b^2 - a^2 \right] = \frac{a+b}{2}.
\]

**Expo(\(\lambda\))**

The exponential distribution with parameter \(\lambda > 0\) is defined by 
\[
f_X(x) = \lambda e^{-\lambda x} \cdot 1(x \geq 0)
\]
\[
F_X(x) = \begin{cases} 
0, & \text{if } x < 0 \\
1 - e^{-\lambda x}, & \text{if } x \geq 0.
\end{cases}
\]

Recall that 
\[
Pr[X \in (i\delta, (i+1)\delta)] = f_X(i\delta)\delta.
\]
Thus, 
\[
E[X] = \sum_{i=-\infty}^{\infty} (i\delta)Pr[i\delta < X \leq (i+1)\delta] = \sum_{i=-\infty}^{\infty} (i\delta)f_X(i\delta)\delta = \int_{-\infty}^{\infty} x f_X(x)\,dx.
\]

**Definition** The expectation, \(E[X]\) of a continuous random variable is defined as 
\[
E[X] = \int_{-\infty}^{\infty} x f(x)\,dx.
\]

**Expectation: dartboard.**

Example: distance from center on radius 1 dartboard. Recall: 
\[
f_Y(y) = 2y 1\{0 \leq y \leq 1\}.
\]
Hence, 
\[
E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_0^1 y 2y\,dy = \frac{2}{3}.
\]
Try whole process for general radius. What do you get?

**Expectation: Exponential.**

Let \(X = \text{Expo}(\lambda)\). Then, 
\[
E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^{\infty} x \lambda e^{-\lambda x}\,dx = -\int_0^{\infty} xde^{-\lambda x} = -\left[ \frac{xe^{-\lambda x}}{\lambda} \right]_0^{\infty} - \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x}\,dx = -\frac{1}{\lambda} - \frac{1}{\lambda} = -\frac{1}{\lambda}.
\]
\(^{(*)}\) We used the integration by parts formula: 
\[
\int_a^b f(x)g'(x)\,dx = [f(x)g(x)]_a^b - \int_a^b g(x)df(x),
\]
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 $[0, \ell]$.

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

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

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

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

---

**Summary**

**Conditional Expectation, Continuous Probability**

1. $E[Y|X] := \sum_{y} y \Pr(Y = y|X = x)$.
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: $E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx$.