Probability Theory

Instructor: Jodin Morey   moreyj@lemoyne.edu

Previous Lecture

- Covariance: $E(XY) - E(X)E(Y)$
- Covariance Properties
- Correlation: $\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

9 - Conditional Expectation

Why? Allows us to:

- **Decompose** complicated expectation problems into a number of easier ones.
- **Estimate** unknowns based on whatever evidence is currently available.

There are two types:

§9.1 – Given rv Y and event A, we have: $E(Y|A)$. (this is just a number)

§9.2 – Given rv Y and the result of another rv X, we have: $E(Y|X)$. (this is a rv!)

§9.1 - Conditional Expectation Given an Event

Discrete: Recall that $E(Y)$ of discrete Y is a **weighted** average of the Y values: $E(Y) = \sum_{i=1}^{\infty} y_i P(Y = y_i)$.

Def (**Conditional Expectation Given a Discrete Event**): After learning that event A occurred, we update the **weights**.
$E(Y|A) = \sum_{i=1}^{\infty} y_i P(Y = y_i|A)$, where the sum is over the support of Y.

Continuous: Similarly, recall that if Y is cont, then $E(Y)$ is also a weighted average of Y values: $E(Y) = \int_{-\infty}^{\infty} yf(y)dy$.

Def (**Conditional Expectation Given a Cont Event**): If we learn that A occurred, we again update the weights:
$E(Y) = \int_{-\infty}^{\infty} yf(y|A)dy$.

The conditional PF $f(y|A)$ can often be obtained as the derivative of the conditional CDF $F(y|A) = P(Y \leq y|A)$.

The conditional PF can also be computed by a hybrid version of Bayes' rule: $f(y|A) = \frac{P(A|Y=y)f(y)}{P(A)}$.

Law of Total Expectation

Let $A_1, \ldots, A_n$ be a partition of a sample space, with $P(A_i) > 0$ for all i, and let B be an event.

Recall **Law of Total Prob**: $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$.

Similarly for rv Y, we have:

Thm (**Law of Total Expectation**): $E(Y) = \sum_{i=1}^{n} E(Y|A_i)P(A_i)$.

---

§9.2 - Conditional Expectation given a Rv

$E(Y|X)$ is our best prediction of Y, assuming we get to know X.

Let's first understand $E(Y|X = x)$.
Since $X = x$ is an event, $E(Y|X = x)$ is just the conditional expectation of Y given this event (as seen above, this is just a number).

Discrete: If Y is discrete, in our calculation for $E(Y|X = x)$ we use the conditional $P(Y = y|X = x)$ in place of the unconditional $P(Y = y)$. So, $E(Y|X = x) = \Sigma_y yP(Y = y|X = x)$.

Continuous: Analogously, if Y is continuous, for $E(Y|X = x)$ we use the conditional $f_{Y|X}(y|x)$ in place of the unconditional $f(y)$. So, $E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$.

Notice that because we sum or integrate over y, $E(Y|X = x)$ is a function of x only. We can call this $g(x) := E(Y|X = x)$. If we do, we can then define...

Def (**Conditional Expectation given an Rv**): Let $g(x) := E(Y|X = x)$. Then the conditional expectation of Y given X, denoted $E(Y|X)$, is the *rv* $g(X)$.
In other words, after the random process, X crystallizes into x, and $E(Y|X)$ crystallizes into $g(x)$.

Ex (**Conditional Expectation**): Let cont X and Y have joint PF: $f(x,y) = \begin{cases} -7y^3 & \text{for } 0 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$

a. Find $E(X|Y)$.

$E(X|Y) = \int_0^y x f_{X|Y}(x|y)dx$.

So we must first find the conditional distribution $f_{X|Y}(x|y)$.

$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$.

So we must first find the marginal distribution $f_Y(y)$.

$f_Y(y) = \int_0^y -7y^3 dx = -7y^3[x]_0^y = -7y^4.$

So, $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(x)} = \frac{-7y^3}{-7y^4} = \frac{1}{y} = \frac{1}{y}$, and

$$E(X|Y) = \int_0^y x f_{X|Y}(x|y)dx = \frac{1}{y}\int_0^y x dx = \frac{1}{2y}[x^2]_0^y = \frac{y^2}{2y} = \frac{y}{2}.$$

Activity 17

Adam and Eve's Laws

The next two thms are picked out of §9.3 and 9.5 (even though we're not covering these sections more fully).
Adam's law connects conditional to unconditional expectation.

Thm (Adam's Law): For any X and Y, we have $E(E(Y|X)) = E(Y)$.

Proof (discrete case). Let $g(X) := E(Y|X)$. Then:

$$E(g(X)) = \Sigma_x g(x)P(X = x) \qquad \text{(LOTUS)}$$

$$= \Sigma_x(\Sigma_y yP(Y = y|X = x))P(X = x) \qquad \text{(def of cond expectation)}$$

$$= \Sigma_x\Sigma_y yP(X = x)P(Y = y|X = x)$$

$$= \Sigma_y y\Sigma_x P(Y = y|X = x)P(X = x) \qquad \text{(swap the order of summation)}$$

$$= \Sigma_y yP(Y = y) = E(Y). \qquad \text{(LOTP and def of expectation)} \qquad \blacksquare$$

A companion result is Eve's law, which relates conditional to unconditional variance.

Thm (Eve's Law): For any X and Y, $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$. (notice *EVVE*)

Proof. Let $g(X) := E(Y|X)$: By Adam's law, $E(g(X)) = E(Y)$. So,

$$E(Var(Y|X)) = E(E(Y^2|X) - g(X)^2) \qquad \text{(expanding variance)}$$

$$= E(Y^2) - E(g(X)^2), \qquad \text{(linearity and Adam's law)}$$

$$Var(E(Y|X)) = E(g(X)^2) - (E(g(X)))^2 \qquad \text{(expanding variance)}$$

$$= E(g(X)^2) - (EY)^2. \qquad \text{(Adam's law)}$$

So, $E(Var(Y|X)) + Var(E(Y|X)) = E(Y^2) - E(g(X)^2) + E(g(X)^2) - (EY)^2$

$$= E(Y^2) - (EY)^2 = Var(Y). \quad \blacksquare$$

Eve's law Intuition:

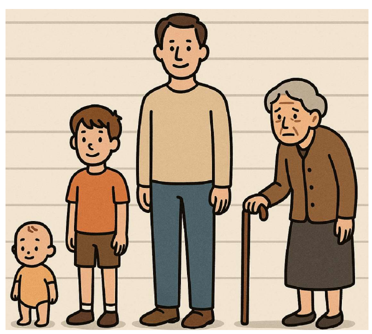
To visualize Eve's law, imagine a population where each person has age X and height Y.

We divide this population ino subpopulations, one for each age X.

Then, there are two things contributing to variation in heights.
Within each age group, people have different heights: **within-group variation**, $E(Var(Y|X))$.

Within-group variation: 30yr old Heights

Across age groups, the average heights are different: **between-group variation**, $Var(E(Y|X))$.



Between-group variation: Various Aged Heights

Eve's law says total variance of Y is the sum of these two sources of variation.

To predict height based on age alone, ideally everyone within an age group would had exactly the same height, while different age groups had different heights. Then, given someone's age, we would be able to predict their height perfectly.

No within-group variation

So, the ideal situation for prediction is to have no within-group variation in height, since the within-group variation cannot be explained by age differences.

Thus, within-group variation is called **unexplained variation**, and between-group variation is called **explained variation**.
Eve's law says that the overall variance of Y is **the sum of unexplained and explained variation**.

Example Problems

Adam and Eve's laws allow us to find the mean and variance of complicated rvs, especially in situations that involve **multiple levels of randomness**.

Ex (**Random Sum**): A store receives N customers in a day, where N is a rv w/finite mean and variance.
Let X_j be amount spent by the jth customer at the store. Assume that each X_j has mean μ and variance σ^2, and that N and all the X_j are indep of one another. Find mean and variance of the random sum $X = \sum_{j=1}^{N} X_j$ (which is the store's total revenue in a day). Give them in terms of μ, σ^2, $E(N)$, and $Var(N)$.

Solution: Since X is a sum, our first impulse might be to claim "$E(X) = N\mu$ by linearity".

This is a category error since $E(X)$ is a number and N is a rv (RHS and LHS of equation aren't the same type of things)

The key is that X is not merely a sum, but a random sum; the # of terms we're adding up is *itself* random, whereas linearity applies to sums with a fixed # of terms.

Alternate strategy: we *wish* to treat N as a constant, so we can use linearity. So let's condition on N.

By linearity of conditional expectation,

$$E(X|N) = E\left(\left(\sum_{j=1}^{N} X_j\right)|N\right) = \sum_{j=1}^{N} E(X_j|N) = \sum_{j=1}^{N} E(X_j) = N\mu.$$

We used the independence of the X_j and N to assert $E(X_j|N) = E(X_j)$ for all j.

Note that the statement "$E(X|N) = N$ " is not a category error because both sides of the equality are rvs that are functions of N.

Finally, by Adam's law, $E(X) = E(E(X|N)) = E(N\mu) = \mu E(N)$.
The average total revenue is the average amount spent per customer, multiplied by the average # of customers.

For $Var(X)$, we again condition on N to get $Var(X|N)$:

$$Var(X|N) = Var\left(\left(\sum_{j=1}^{N} X_j\right)|N\right) = \sum_{j=1}^{N} Var(X_j|N) = \sum_{j=1}^{N} Var(X_j) = N\sigma^2.$$

Eve's law then tells us how to obtain the unconditional variance of X:

$$Var(X) = E(Var(X|N)) + Var(E(X|N)) = E(N\sigma^2) + Var(N\mu) = \sigma^2 E(N) + \mu^2 Var(N). \quad \square$$

Ex (**Adam&Eve Dice**): Let Y be the outcome of rolling a fair six-sided die. Define $X = Y + Z$, where $Z\,|\,(Y = y) \sim Unif(-y, y)$.
That is, given that the die shows $Y = y$, then Z is uniformly distr from $-y$ to y. Find: $E(X)$ and $Var(X)$.

Solution: I'd like to say: $E(X) = E(Y+Z)$, and then use linearity. However, technically the result would be a rv, not a number since $E(Z)$ depends (albeit trivially) on Y.

So, instead we first calculate $E(X\,|\,Y)$, and then apply Adam's law: $E(X) = E[E(X\,|\,Y)]$.

Since $X = Y+Z$, we have: $E(X\,|\,Y) = E(Y+Z\,|\,Y)$

$$= E(Y\,|\,Y) + E(Z\,|\,Y) = Y+0 = Y$$

(because the mean of a symmetric uniform distr over $[-y,y]$ is 0.)

Therefore, by Adams law: $E(X) = E[E(X\,|\,Y)] = E(Y)$.

And, since $Y \sim Unif\{1,2,3,4,5,6\}$, we compute: $E(Y) = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}$. So, $E(X) = \frac{7}{2}$.

Regarding Variance, Eve's Law states: $Var(X) = E[Var(X\,|\,Y)] + Var(E(X\,|\,Y))$.

For the first term, we calculate, $Var(X\,|\,Y) = Var(Y+Z\,|\,Y)$

$$= Var(Y\,|\,Y) + Var(Z\,|\,Y)$$

(Y is a constant in this context, since we are conditioning on Y, so Z and Y are indep)

$$= 0 + Var(Z\,|\,Y) = Var(Z\,|\,Y).$$

Recall the variance of $Unif(a,b)$ is: $\frac{(b-a)^2}{12}$.

So, since $Z\,|\,(Y=y) \sim Unif(-y,y)$, we have $Var(Z\,|\,Y=y) = \frac{(2y)^2}{12} = \frac{4y^2}{12} = \frac{y^2}{3}$.

Thus: $E[Var(X\,|\,Y)] = E\left(\frac{Y^2}{3}\right) = \frac{1}{3}E(Y^2)$.

We compute: $E(Y^2) = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}$, (LOTUS)

Thus, $E[Var(X\,|\,Y)] = \frac{1}{3}\frac{91}{6} = \frac{91}{18}$.

For the second Eve term, $Var(E(X\,|\,Y))$, note we've already found: $E(X\,|\,Y) = Y$, therefore $Var(E(X\,|\,Y)) = Var(Y)$.

And $Var(Y) = E(Y^2) - (E(Y))^2 = \frac{91}{6} - (\frac{7}{2})^2 = \frac{35}{12}$. (we computed these earlier)

Now we compute Eve's total variance: $Var(X) = E[Var(X\,|\,Y)] + Var(E(X\,|\,Y)) = \frac{91}{18} + \frac{35}{12} = \frac{287}{36}$.

Activity 18

What did we learn?

- Conditional Expectation Given an Event: scalar $E(Y|A)$
- Conditional Expectation Given a rv: rv $E(Y|X)$
- Adam's Law
- Eve's Law