MA 6.101
Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Conditioning with random variables

Today's class

- Conditioning X on an event $A \in \mathcal{F}$.
- Conditional Expectation $E[X|A]$.
- Conditioning X with disjoint partitions $\{A_i\}$ of Ω.
- Conditioning X on an event $\{X \in A\} \in \mathcal{F}'$
- Conditioning X on another random variable Y.
- Conditional expectation $E[X|Y = y]$.

A new running example

- Pick 2 integers from $\{1, 2, 3\}$ without replacement.
- $\Omega = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$
- $\mathbb{P}\{\omega\} = \frac{1}{6}$ for all $\omega \in \Omega$.
- Denote them by random variables X and Y.
- For $\omega = (1, 3)$ $X(\omega) = 1$ and $Y(\omega) = 3$.
- Write down their joint PMF $p_{X,Y}(x, y)$.
- Write down their marginal PMFs p_X and p_Y?
- What is $E[X], E[Y]$ and $E[XY]$?

Remember Conditional probability?

- Given/If dice rolls odd, what is the probability that the outcome is 1?
- Given/If $\bar{\omega} \in [0, 0.5]$ what is the probability that $\bar{\omega} \in [0, 0.25]$?
- The conditional probability of event B given event A is defined as $\mathbb{P}(B/A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ when $\mathbb{P}(A) > 0$.

Conditioning on an event A

- Consider a discrete r.v. X with pmf $p_X(x)$. Suppose an event A has happened where $A \in \mathcal{F}$.
- Consider event $\{\omega \in \Omega : X(\omega) = x\}$. We will use shorthand $\{X = x\}$.
- What is $\mathbb{P}(X = x|A)$? $\mathbb{P}(X = x|A) = \frac{\mathbb{P}(\{X=x\}\cap A)}{\mathbb{P}(A)}$.

$$p_{X|A}(x) := \frac{\mathbb{P}(\{X=x\}\cap A)}{\mathbb{P}(A)}.$$

- $p_{X|A}(x)$ denotes the conditional PMF of X under event A.
- In the running example say A is the event that the first number is odd and second is even. $A = \{(1,2),(3,2)\}$. Compute $p_{X|A}(\cdot)$.
- How do we know that it is consistent, i.e., $\sum_x p_{X|A}(x) = 1$?

Consistency of conditional PMF

$$\sum_x p_{X|A}(x) = 1.$$

Proof:

- $\sum_x p_{X|A}(x) = \sum_x \frac{\mathbb{P}(\{X=x\}\cap A)}{\mathbb{P}(A)}$
- $\{\omega \in \Omega : X(\omega) = x\}$ are disjoint sets for different x.
- From theorem of total probability, this implies that $\{X = x\} \cap A$ are disjoint sets for all x.
- $\sum_x p_{X|A}(x) = \frac{\mathbb{P}(\bigcup_x \{\{X=x\}\cap A)\}}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$ $\square$

Another Example

- Lets X denote the outcome of a dice.
- Let A denote the event that the roll is odd.
- What is $p_{X|A}(x)$?
- Given that event A has happened, what is the average value of the dice, i.e., $E[X|A]$?

$$E[X/A] = \sum_x x p_{X|A}(x).$$

Using LOTUS,

$$E[g(X)/A] = \sum_x g(x) p_{X|A}(x).$$

Today's class

- Conditioning X on an event $A \in \mathcal{F}$.
- Conditional Expectation $E[X|A]$.
- Conditioning X with disjoint partitions $\{A_i\}$ of Ω.
- Conditioning X on an event $\{X \in A\} \in \mathcal{F}'$
- Conditioning X on another random variable Y.
- Conditional expectation $E[X|Y = y]$.

Conditioning with disjoint partitions

- Now let $\{A_i, i = 1, 2, \ldots, n\}$ be a disjoint partition of Ω.
- Prove the following using law of total probability

$$p_X(x) = \sum_{i=1}^{n} \mathbb{P}(A_i) p_{X|A_i}(x)$$

Proof:

- $\sum_{i=1}^{n} \mathbb{P}(A_i) \frac{\mathbb{P}(\{X=x\} \cap A_i)}{\mathbb{P}(A_i)} = \sum_{i=1}^{n} \mathbb{P}(\{X = x\} \cap A_i) = \mathbb{P}(\{X = x\}).$ $\square$
- The last equality follows from the law of total probability.
- An important consequence is the following.

$$E[X] = \sum_{i=1}^{n} \mathbb{P}(A_i) E[X|A_i]$$

Today's class

- Conditioning X on an event $A \in \mathcal{F}$.
- Conditional Expectation $E[X|A]$.
- Conditioning X with disjoint partitions $\{A_i\}$ of Ω.
- Conditioning X on an event $\{X \in A\} \in \mathcal{F}'$
- Conditioning X on another random variable Y.
- Conditional expectation $E[X|Y = y]$.

Conditioning on event $X \in A$

- Consider a discrete r.v. X with pmf $p_X(x)$. Suppose an event $X \in A$ has happened where $A \in \mathcal{F}'$.
- $X \in A = \{\omega \in \Omega : X(\omega) \in A\}$ and $\mathbb{P}\{X \in A\} = \sum_{x \in A} p_X(x)$.
- We will use the same notation $p_{X|A}(x) := \frac{\mathbb{P}(\{X=x\} \cap \{X \in A\})}{\mathbb{P}(X \in A)}$.
- If $x \notin A$, we have $p_{X|A}(x) = 0$.
- Otherwise (when $x \in A$,), we have $p_{X|A}(x) = \frac{p_X(x)}{\mathbb{P}(X \in A)}$.
- Running example: Suppose we are given $X \in A$ where $A = \{2, 3\}$. What is $p_{X|A}(x)$?

Revisiting Geometric random variable

- Let N be a geometric random variable with parameter p.
- Its pmf is $p_N(k) = (1-p)^{k-1}p$.
- Suppose we are given the event $A := N > n$. $P(A) = (1-p)^n$.
- What is $p_{N|A}(k)$?
- For $k > n$, $p_{N|A}(k) = \frac{P\{(N>n) \cap N=k\}}{P(N>n)} = (1-p)^{k-1-n}p$. For $k \leq n$, we have $p_{N|A}(k) = 0$.

Memoryless property of Geometric random variable

- What is $P(N > n + m | N > n)$?
- $P(N > n + m | N > n) = \frac{P(N > n+m)}{P(N > n)} = (1-p)^m = P(N > m)$.
- If N denotes number of tosses till you first get a head, and having already tossed more than n times, the probability of having to toss more than $n + m$ is same as starting the experiment (forgetting that you have already tossed more than n times) fresh and having to toss more than m times.
- How much you have tossed till now has no bearing on how much you will be required to toss.

$$\boxed{P(N > n + m | N > n) = P(N > m) \text{ (Memoryless property).}}$$

HW: Find $E[N|A]$ where event $A = \{N > n\}$ and $n > 0$.

Today's class

- Conditioning X on an event $A \in \mathcal{F}$.
- Conditional Expectation $E[X|A]$.
- Conditioning X with disjoint partitions $\{A_i\}$ of Ω.
- Conditioning X on an event $\{X \in A\} \in \mathcal{F}'$
- Conditioning X on another random variable Y.
- Conditional expectation $E[X|Y = y]$.
- Law of iterated expectation $E[X|Y]$
- Bayes rule revisited
- Sums of random variables.

Conditioning X on random variable Y

- Consider a discrete r.v's X and Y with joint pmfs $p_{XY}(x,y)$ and with marginal pmf $p_X(x)$ and $p_Y(y)$.
- Suppose an event $A : \{Y = y\}$ has happened and we are interested in the probability that $X = x$ given $Y = y$.
- This conditional pmf is denoted by $p_{X|Y}(x|y)$.
- In fact, $p_{X|Y}(x|y) := \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$.

$$\boxed{p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y)}$$

- This is essentially same as $P(A \cap B) = P(A|B)P(B)$
- Is $p_{X|Y}(x|y)$ consistent?
- $\sum_x p_{X|Y}(x|y) = \sum_x \frac{p_{X,Y}(x,y)}{p_Y(y)} = 1$.

Conditioning X on random variable Y

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y)$$

- Now summing on both sides over y, we have

$$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$$

- Similarly from $p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x)$, summing on both sides over x, we have

$$p_Y(y) = \sum_x p_{Y|X}(y|x)p_X(x)$$

- Notice similarity to the law of total probabiltiy.
  $P(A) = \sum_i P(A|B_i)P(B_i)$.

Conditional expectation $E[X|Y=y]$

It is easy to guess that

$$E[X|Y=y] := \sum_x x p_{X|Y}(x|y)$$
$$E[Y|X=x] := \sum_y y p_{Y|X}(y|x)$$

Can you write $E[X]$ in terms of $E[X|Y=y]$?

$$E[X] = \sum_y p_Y(y) E[X|Y=y]$$

Proof:
$$\begin{aligned}
\sum_y p_Y(y) E[X|Y=y] &= \sum_y p_Y(y) \sum_x x p_{X|Y}(x|y) \\
&= \sum_x \sum_y x p_{X,Y}(x,y) \\
&= \sum_x x p_X(x) \\
&= E[X]
\end{aligned}$$

Summary

$$p_{X|A}(x) := \frac{\mathbb{P}(\{X=x\}\cap A)}{\mathbb{P}(A)}.$$

$$E[X/A] = \sum_x x p_{X|A}(x).$$

$$p_X(x) = \sum_{i=1}^n \mathbb{P}(A_i) p_{X|A_i}(x)$$

$$E[X] = \sum_{i=1}^n \mathbb{P}(A_i) E[X|A_i]$$

$$p_{X,Y}(x,y) = p_{X|Y}(x|y) p_Y(y)$$

$$p_X(x) = \sum_y p_{X|Y}(x|y) p_Y(y)$$

$$E[X|Y=y] := \sum_x x p_{X|Y}(x|y)$$

$$E[X] = \sum_y p_Y(y) E[X|Y=y]$$

How about all this for continuous X & Y?

$$p_{X|A}(x) := \frac{\mathbb{P}(\{X=x\}\cap A)}{\mathbb{P}(A)}.$$

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y)$$

$$E[X/A] = \sum_x x p_{X|A}(x).$$

$$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$$

$$p_X(x) = \sum_{i=1}^n \mathbb{P}(A_i)p_{X|A_i}(x)$$

$$E[X|Y=y] := \sum_x x p_{X|Y}(x|y)$$

$$E[X] = \sum_{i=1}^n \mathbb{P}(A_i)E[X|A_i]$$

$$E[X] = \sum_y p_Y(y)E[X|Y=y]$$

How about all this for continuous X & Y?

$\int_{x \in B} f_{X|A}(x) dx = \mathbb{P}(X \in B|A).$

$E[X/A] = \sum_x x p_{X|A}(x).$

$p_X(x) = \sum_{i=1}^n \mathbb{P}(A_i) p_{X|A_i}(x)$

$E[X] = \sum_{i=1}^n \mathbb{P}(A_i) E[X|A_i]$

$p_{X,Y}(x,y) = p_{X|Y}(x|y) p_Y(y)$

$p_X(x) = \sum_y p_{X|Y}(x|y) p_Y(y)$

$E[X|Y=y] := \sum_x x p_{X|Y}(x|y)$

$E[X] = \sum_y p_Y(y) E[X|Y=y]$

How about all this for continuous X & Y?

$$\int_{x\in B} f_{X|A}(x) = \mathbb{P}(X \in B|A).$$

$$E[X/A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$

$$p_X(x) = \sum_{i=1}^{n} \mathbb{P}(A_i) p_{X|A_i}(x)$$

$$E[X] = \sum_{i=1}^{n} \mathbb{P}(A_i) E[X|A_i]$$

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y)$$

$$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$$

$$E[X|Y=y] := \sum_x x p_{X|Y}(x|y)$$

$$E[X] = \sum_y p_Y(y)E[X|Y=y]$$

How about all this for continuous X & Y?

$$\int_{x\in B} f_{X|A}(x) = \mathbb{P}(X \in B|A).$$

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y)$$

$$E[X/A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$

$$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$$

$$f_X(x) = \sum_{i=1}^n \mathbb{P}(A_i) f_{X|A_i}(x)$$

$$E[X|Y=y] := \sum_x x p_{X|Y}(x|y)$$

$$E[X] = \sum_{i=1}^n \mathbb{P}(A_i)E[X|A_i]$$

$$E[X] = \sum_y p_Y(y)E[X|Y=y]$$

How about all this for continuous X & Y?

$$\int_{x\in B} f_{X|A}(x) = \mathbb{P}(X \in B|A).$$

$$f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y)$$

$$E[X/A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$

$$p_X(x) = \sum_y p_{X|Y}(x|y) p_Y(y)$$

$$f_X(x) = \sum_{i=1}^n \mathbb{P}(A_i) f_{X|A_i}(x)$$

$$E[X|Y=y] := \sum_x x p_{X|Y}(x|y)$$

$$E[X] = \sum_{i=1}^n \mathbb{P}(A_i) E[X|A_i]$$

$$E[X] = \sum_y p_Y(y) E[X|Y=y]$$

How about all this for continuous X & Y?

$$\int_{x\in B} f_{X|A}(x) = \mathbb{P}(X \in B|A).$$

$$E[X/A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$

$$f_X(x) = \sum_{i=1}^{n} \mathbb{P}(A_i) f_{X|A_i}(x)$$

$$E[X] = \sum_{i=1}^{n} \mathbb{P}(A_i) E[X|A_i]$$

$$f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y)$$

$$f_X(x) = \int_y f X|Y(x|y) f_Y(y) dy$$

$$E[X|Y=y] = \int_x x f_{X|Y}(x|y) dx$$

$$E[X] = \int_y E[X|Y=y] f_Y(y) dy$$

Conditional expectation $E[X|Y]$

Recall that

$$E[X|Y = y] := \sum_x x p_{X|Y}(x|y)$$

- Is $E[X|Y = y]$ a constant? Is it a function of y?
- $E[X|Y = y]$ is a function of y.
- Now consider $E[X|Y]$. Is it still a function of y ?
- $E[X|Y]$ is a function of Y, say $g(Y)$.
- When Y takes the value y, (this happens with probability $p_Y(y)$) $E[X|Y]$ takea the value $E[X|Y = y]$.
- What is the expectation of $E[X|Y]$?

Conditional expectation $E[X|Y]$

- $g(Y) = E[X|Y]$.
- What is $E[g(Y)] = E[E[X|Y]]$?
- $E[g(Y)] = \sum_y g(y)p_Y(y) = \sum_y E[X|Y = y]p_Y(y)$.
- This implies $E[g(Y)] = E[E[X|Y]] = E[X]$. This is the law of iterated expectation.

$$\boxed{E[E[X|Y]] = E[X]}$$

Conditional expectation $E[X|Y]$ – Example

- Consider $Y = \begin{cases} \lambda_1 \text{ with prob } p \\ \lambda_2 \text{ with prob } 1-p \text{ .} \end{cases}$
- Now consider an exponential random variable X with a random parameter Y.
- What is $E[X]$?
- $E[X] = E[E[X|Y]] \sum_y E[X|Y = y] p_Y(y)$
- We have $X \sim Exp(\lambda_1)$ with probability p when $Y = \lambda_1$.
- Similarly $X \sim Exp(\lambda_2)$ with probability $1-p$ when $Y = \lambda_2$.
- $E[X|Y = \lambda_i] = \frac{1}{\lambda_i}$
- $E[X] = \frac{p}{\lambda_1} + \frac{1-p}{\lambda_2}$.

Conditional expectation $E[X|Y]$ – Example 2

- Consider $Y = X_1 + X_2 + \ldots X_N$ where N is a positive integer valued r.v. with PMF $p_N(\cdot)$ and $X_i's$ are independent and identically distributed (i.i.d) with mean $E[X]$.
- What is $E[Y]$? Use $E[Y] = E[E[Y|N]]$.
- What is $E[Y|N = n]$?
- $E[Y|N = n] = E[X_1 + X_2 + \ldots X_n] = nE[X]$.
- This implies $E[Y|N] = NE[X]$.
- Now $E[Y] = E[E[Y|N]] = E[NE[X]] = E[X]E[N]$.
- What is $Var(Y)$? (section 4.5)

Bayes Rule revisited

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)}$$

For discrete random variables X and Y

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$$

For continuous random variables X and Y

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Sums of independent random variable

- Consider $Z = X + Y$. What is the pdf of Z when X and Y?
- What is $p_Z(z)$ or $f_Z(z)$?
- $p_Z(z) = \sum_{(x,y):x+y=z} p_{X,Y}(x,y)$
- $f_Z(z) = \int_{(x,y):x+y=z} f_{X,Y}(x,y)$.
- Since X and Y are independent $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ and $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. This gives us

Convolution formula

$$p_Z(z) = \sum_x p_X(x)p_Y(z-x)$$
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

HW: What if X and Y are not independent?

MGF of Sums of independent random variable

- Consider $Z = X + Y$. What is the pdf of Z when X and Y?
- Let $M_X(t)$ and $M_Y(t)$ be their MGF's. What is $M_Z(t)$?
- $M_Z(t) = E[e^{Zt}] = E[e^{(X+Y)t}]$.
- $M_Z(t) = E[e^{Xt}.e^{Yt}]$.
- If X and Y are independent, $E[XY] = E[X]E[Y]$ and $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.
- $M_Z(t) = E[e^{Xt}].E[e^{Yt}]$.

$$\boxed{M_Z(t) = M_X(t)M_Y(t).}$$

MGF of Sums of independent random variable

- Consider $Z = X + Y$. What is the MGF of Z when X and Y?

$$M_Z(t) = M_X(t)M_Y(t).$$

- What about $M_Z(t)$ when $Z = X_1 + X_2 + \ldots X_n$ and X_i are iid.?
- $M_Z(t) = (M_X(t))^n$.
- What about $M_Z(t)$ when $Z = X_1 + X_2 + \ldots X_N$ where N is a positive discrete random variable? section 4.5