ECSE 2500
Lec 9
Feb 23th

Topic: Conditional PMF
Conditional Expected Values
Conditional Variance

Remember that the definition of Probability Mass function (PMF)

$$P_X(x_i) = P(X = x_i) = P(s \in \underset{\uparrow \text{Sample space}}{S} \mid X(s) = x_i)$$

We now combine this with the notion of conditional probability to talk about

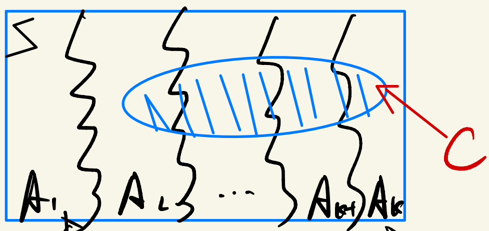
Conditional PMF

$$P_X(x_i \mid C) = P(X = x_i \mid C)$$

e.g., Probability that packet arrives in 5 seconds given that we have already waited for 5 seconds.

Alternatively, we can think of conditional PMF

$$P_X(x_i|C) = \frac{P[s\in S \mid X(s)=x_i \cap s\in C]}{P(C)} \Leftarrow P(s\in C)$$



$P(s\in S \mid X(s)=x_1)$
$=P(A_1)$

$P(s\in S \mid X(s)=x_k)$
$=P(A_k)$

Notice: We implicitly assume that $P(C)>0$ and $C\neq\phi$. Otherwise, $P_X(x_i|C)=0$.

Example: What's the probability we got 5 heads on 5 biased coin tosses, given that the first two tosses were heads?

X = # of heads in 5 coin tosses

$P(\text{head})=p$, $P(\text{tail})=1-p$, $p\neq\frac{1}{2}$.

$$P(X=5 \mid \text{first 2 heads}) = \frac{P(HHHHH \cap HHXXX)}{P(HHXXX)}$$

$$= \frac{P(HHHHH)}{P(HHXXX)}$$

$$= \frac{P(H)^5}{P(H)\cdot P(H)} = p^3$$

If coin is fair, $p = \frac{1}{2}$,

$$P(X=5 \mid \text{first 2 heads}) = \frac{1}{8}$$

$$P(X=5) = \frac{1}{32} \ll \frac{1}{8}$$

Notice: The notion $P_X(x_i \mid C)$ is still a probability mass function – satisfying all properties/rules of PMFs:

1) $P_X(x_i \mid C) \geq 0$;

2) $\sum_{i=1}^{N} P_X(x_i \mid C) = 1$, where N is the number of possible values X can take;

3) Total probability theorem:

If $\{B_1, B_2, \cdots, B_M\}$ is a partition of the original sample space, then

$$P_X(x_i) = \sum_{i=1}^{M} P_X(x_i \mid B_i)\, P(B_i)$$

Example: Waiting for a bus. Suppose that the

X : # of minutes we must wait is uniform,

$$P_X(k) = \begin{cases} \frac{1}{20}, & k = 1, 2, \cdots, 20 \\ 0, & \text{otherwise} \end{cases}$$

Given we have waited for 8 minutes, what is the conditional PMF of waiting time?

$$P_X(k \mid X > 8) = \frac{P(X = k \text{ and } X > 8)}{P(X > 8)}$$

$$= \frac{\frac{1}{20}}{\frac{12}{20}}$$

$$= \frac{1}{12}$$

Notice: In the above calculation, we made the assumption that $k > 9$;

If $k \leq 8$, $P(X = k \text{ and } X > 8) = 0$,

$$P_X(k \mid X > 8) = 0$$

Example: Let

X = # of Additional Years a randomly chosen 70 years old person can live.

Event H: The person has high blood pressure;

X will be a Geometric RV with parameter $p=0.1$

Event N: The person has Normal Blood pressure;

X will be a Geometric RV with parameter $p=0.05$.

$\boxed{\text{Recall}}$ $P_X(k) = (1-p)^{k-1}p, \quad k=1,2,\cdots\infty$

$$P_{X|H}(k) = P_X(k|H) = (1-0.1)^{k-1}\cdot(0.1)$$

$$P_{X|N}(k) = P_X(k|N) = (1-0.05)^{k-1}\cdot(0.05)$$

If 40% of 70 years old person has high blood pressure, what is the PMF of X?

$$P_X(k) = P_{X|H}(k)\cdot P(H) + P_{X|N}(k)\cdot P(N)$$

$$= (0.9)^{k-1}\cdot(0.1)\cdot(0.4) + (0.95)^{k-1}\cdot(0.05)\cdot(0.6)$$

⇒ We can also define conditional expected values:

$$E[X|B] = \sum_{x_i \in S_X} x_i \, P_X(x_i|B)$$

If we know the conditional expected value of X, we can also compute

$$E[X] = \sum_{x_i \in S_X} x_i \, P_X(x_i)$$

Total probability theorem

$$= \sum_{x_i \in S_X} x_i \sum_{j=1}^{M} P_X(x_i|B_j)\, P(B_j)$$

$$= \sum_{j=1}^{M} \underbrace{\sum_{x_i \in S_X} x_i \, P_X(x_i|B_j)}\, P(B_j)$$

$$= \sum_{j=1}^{M} E[X|B_j]\, P(B_j)$$

The above relation is convenient since we are often given the latter and we don't need to explicitly create the PMF.

Caveat: Variance does not follow the above relation in the sense that

$$Var(X) = Var(X|H) \cdot P(H) + Var(X|N) P(N)$$

What is conditional variance? How to calculate it?

$$Var(X|C) = E[X^2|C] - (E[X|C])^2$$

$$\boxed{Var(X) = E[X^2] - (E[X])^2}$$

☐ Memoryless property of Geometric Random Variable

Say X is a Geometric Random Variable with parameter p, i.e.,

$$P_X(k) = (1-p)^{k-1} \cdot p, \quad k = 1, 2, \ldots \infty$$

What is $P(X > j + j_0 \mid X > j)$?

$$= P(X > j_0)$$

To show the above memoryless property,

$$P(X>j) = \sum_{k=j+1}^{\infty} (1-p)^{k-1} \cdot p$$

$$= (1-p)^{j} \sum_{k=1}^{\infty} (1-p)^{k-1} \cdot p$$

$$= (1-p)^{j}$$

Likewise, we can compute

$$P(X>j+j_0) = \sum_{k=j+j_0+1}^{\infty} (1-p)^{k-1} \cdot p$$

$$= (1-p)^{j_0+j}$$

Therefore, we have

$$P(X>j+j_0 \mid X>j) = \frac{P(X>j+j_0)}{P(X>j)}$$

$$= \frac{(1-p)^{j_0+j}}{(1-p)^{j}}$$

$$= (1-p)^{j_0} = P(X>j_0)$$

Interpretation of Memoryless Property:

If success hasn't occurred in the first j trials, the probability of rerunning j_0 more trials is the same as running j_0 trials in the first place

(i.e., coin does not "memorize" previous history, if we fail, it is the same as starting again)