STAT 430/510 Probability

Hui Nie

Lecture 19

June 30th, 2009

Review

- Covariance and Correlation
- Independence vs Uncorrelated
- Conditional Expectation

Computing Expectations by Conditioning

- $E[X] = E[E[X|Y]]$
- If Y is a discrete random variable, $E[X] = \sum_y E[X|Y = y]P(Y = y)$
- If Y is a continuous random variable, $E[X] = \int_{-\infty}^{\infty} E[X|Y = y]f_Y(y)dy$

Covariance

- A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

Example: Solution

- Let X denote the amount of time until the miner reaches safety, and let Y denote the door he initially chooses.
- $E[X] = E[X|Y = 1]P(Y = 1) + E[X|Y = 2]P(Y = 2) + E[X|Y = 3]P(Y = 3)$
- $E[X|Y = 1] = 3$, $E[X|Y = 2] = 5 + E[X]$, $E[X|Y = 3] = 7 + E[X]$
- $P(Y = 1) = P(Y = 2) = P(Y = 3) = 1/3$
- $E[X] = 15$

Expectation of a Sum of a Random Number of R.V.'s

- Suppose that the number of people entering a department store on a given day is a random variable with mean 50. Suppose further that the amounts of money spent by these customers are independent random variables having a common mean of 8 dollars. Finally, suppose also that the amount of money spent by a customer is also independent of the total number of customers who enter the store. What is the expected amount of money spent in the store on a given day?

Example: Solution

- Let N denote the number of customers that enter the store and X_i be the amount spent by the ith customer.
- $E[\sum_{i=1}^{N} X_i] = E[E[\sum_{i=1}^{N} X_i | N]] = E[N]E[X_1]$
- $50 \times 8 = 400$

Computing Probability by Conditioning

- E is an arbitrary event
- If Y is a discrete random variable, $P(E) = \sum_y P(E|Y = y)P(Y = y)$
- If Y is a continuous random variable, $P(E) = \int_{-\infty}^{\infty} P(E|Y = y)f_Y(y)dy$

Example: The Best-prize problem

- Suppose that we are to be presented with n distinct prizes, in sequence. After being presented with a prize, we must immediately decide whether to accepted or to reject it and consider the next prize. The only information we are given when deciding whether to accept a prize is the relative rank of that prize compared to ones already seen. That is, for instance, when the fifth prize is presented, we learn how it compares with the four prizes we have already seen. Suppose that once a prize is rejected, it is lost, and that our objective is to maximize the probability of obtaining the best prize. Assuming that all $n!$ orderings of the prizes are equally likely, how well can we do?

Example: Solution

- Strategy: Reject the first k prizes and then accepts the first one that is better than all of those first k.
- Let $P_k(best)$ be the probability that the best prize is selected when this strategy is employed and X be the position of the best prize
- $P_k(best) = \sum_{i=1}^{n} P_k(best|X = i)P(X = i) = \frac{1}{n}\sum_{i=1}^{n} P_k(best|X = i)$
- If $i \leq k$, $P_k(best|X = i) = 0$
- If $i > k$, $P_k(best|X = i) =$ $P(\text{best of first i-1 is among the first k}|X = i) = \frac{k}{i-1}$

Properties

$$
\begin{aligned}
P_k(best) &= \frac{k}{n}\sum_{i=k+1}^{n}\frac{1}{i-1} \\
&\approx \frac{k}{n}\int_{k+1}^{n}\frac{1}{x-1}dx \\
&= \frac{k}{n}log(\frac{n-1}{k}) \\
&\approx \frac{k}{n}log(\frac{n}{k})
\end{aligned}
$$

- $P_k(best)$ attains maximum $1/e \approx 0.3679$ at $k^* = n/e$
- Let the first n/e prizes go by and then accept the first one to appear that is better than all of those.

Example

- Let U be a uniform random variable on (0,1), and suppose that the conditional distribution of X, given that $U = p$, is binomial with parameters (n, p). Find the probability mass function of X

$$\begin{aligned} P(X = i) &= \int_0^1 P(X = i|U = p) f_U(p) dp \\ &= \int_0^1 P(X = i|U = p) dp \\ &= \frac{n!}{i!(n-i)!} \int_0^1 p^i (1-p)^{n-i} dp \end{aligned}$$

- $\int_0^1 p^i (1-p)^{n-i} dp = \frac{i!(n-i)!}{(n+1)!}$
- $P(X = i) = \frac{1}{n+1}, i = 0, \cdots, n$

Conditional Variance

- The conditional variance of X given $Y = y$ is defined as

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

- The conditional variance formula:

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

Example: Variance of a Sum of a Random Number of R.V.'s

- Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, and let N be a nonnegative integer-valued random variable that is independent of the sequence X_i, $i \geq 1$. Then,

$$Var(\sum_{i=1}^{N} X_i) = E[N]Var(X) + (E[X])^2 Var(N)$$

Example: Solution

- $E[\sum_{i=1}^{N} X_i | N] = NE[X]$
- $Var(\sum_{i=1}^{N} X_i | N) = NVar(X)$

$$\begin{aligned} Var(\sum_{i=1}^{N} X_i) &= E[Var(\sum_{i=1}^{N} X_i | Y)] + Var(E[\sum_{i=1}^{N} X_i | Y]) \\ &= E[NVar(X)] + Var(NE[X]) \\ &= E[N]Var(X) + (E[X])^2 Var(N) \end{aligned}$$