Conditional Probability
and
Conditional Expectation

Introduction

- One of the most useful concepts in probability theory is that of conditional probability and conditional expectation.
  - In practice, we are often interested in calculating probabilities and expectations when some partial information is available; hence, the desired probabilities and expectations are conditional ones.
  - Secondly, in calculating a desired probability or expectation it is often extremely useful to first "condition" on some appropriate random variable.

- Recall that for any two events E and F, the conditional probability of E given F is defined, as long as $P(F) > 0$, by
  - $P(E|F) = \frac{P(EF)}{P(F)}$

The Discrete Case

- If X and Y are discrete random variables, then it is natural to define the *conditional probability mass function* of X given that $Y = y$, for all values of y such that $P\{Y = y\} > 0$, by
  - $p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{p_{XY}(x, y)}{p_Y(y)}$
- Similarly, the conditional probability distribution function of X given that $Y = y$, for all values of y such that $P\{Y = y\} > 0$, by
  - $F_{X|Y}(x|y) = P(X \le x|Y = y) = \sum_{a \le x} p_{X|Y}(a|y)$
- The conditional expectation of X given that $Y = y$ is defined by
  - $E[X|Y = y] = \sum_x xP(X = x|Y = y) = \sum_x p_{X|Y}(x|y)$
- If X is independent of Y, then
  - $p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X=x,Y=y)}{P(Y=y)} = P(X = x)$

Example 1

- **Example 3.1** Suppose that $p(x, y)$, the joint probability mass function of X and Y, is given by $p(1, 1) = 0.5$, $p(1, 2) = 0.1$, $p(2, 1) = 0.1$, $p(2, 2) = 0.3$. Calculate the conditional probability mass function of X given that $Y = 1$.

  - $p_Y(1) =$

  - $p_{X|Y}(1|1) =$

  - $p_{X|Y}(2|1) =$

Example 2

- **Example 3.2** If X_1 and X_2 are independent binomial random variables with respective parameters (n_1, p) and (n_2, p), calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$.

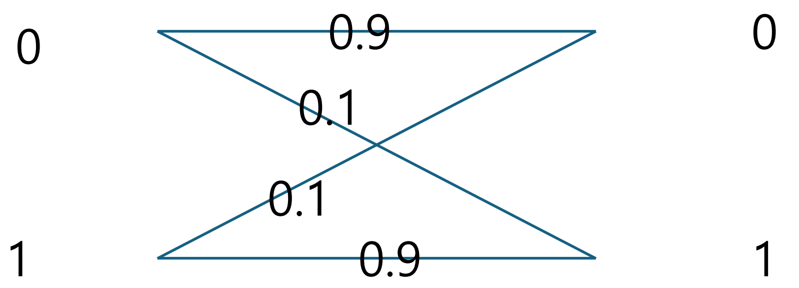
  - Hypergeometric distribution
    - The number of blue balls that are chosen when a sample of m balls is randomly chosen from an urn that contains n_1 blue and n_2 red balls

Example 3

- **Example 3.3** If X and Y are independent Poisson random variables with respective means λ_1 and λ_2, calculate the conditional expected value of X given that $X + Y = n$.

Example 2 of Probability on Lect. 1

- Let A_1 (A_0) and B_1 (B_0) be the event that 1 (0) is sent and the event that 1(0) is received, respectively.
- Assumption
  - $P(A_0) = 0.8$, $P(A_1) = 1 - P(A_0) = 0.2$,
  - The probability of error, i.e., $p=P(B_1|A_0)= P(B_0|A_1)$, is 0.1



- Find
  - The error probability at the receiver
  - The probability that 1 is sent when the receiver decides 1.

Answer: Example 2 on Lect. 1

- Parameters
  - $P(A_0) = 0.8$, $P(A_1) = 1- P(A_0) = 0.2$, $P(B_1|A_0)= P(B_0|A_1) = 0.1$
- The error probability at the receiver
  - P(error)=P$(A_0 B_1)$ + P$(A_1 B_0)$=P$(B_1|A_0)$P(A_0) + P$(B_0|A_1)$P$(A_1) = 0.1$
- The probability that 1 is sent when the receiver decides 1.
  - $P(A_m|B)=\frac{P(B|A_m)P(A_m)}{P(B)} = \frac{P(B|A_m)P(A_m)}{\sum_{n=1}^{N} P(B|A_n)P(A_n)}$
  - $P(A_1|B_1)=\frac{P(B_1|A_1)P(A_1)}{P(B_1)} = \frac{P(B_1|A_1)P(A_1)}{P(B_1|A_1)P(A_1)+P(B_1|A_0)P(A_0)} = \frac{0.9\times 0.2}{0.9\times 0.2+0.1\times 0.8} =$
  0.69

The Continuous Case

- If X and Y have a joint probability density function $f(x, y)$, then the *conditional probability density function* of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by
  - $f_{X|Y}(x|y) = \frac{f(x,y)}{f(y)}$
- To motivate this definition, multiply the left side by dx and the right side by $(dx\ dy)/dy$ to get
  - $f_{X|Y}(x|y)dx = \frac{f(x,y)dxdy}{f(y)dy} \approx \frac{P\{x \le X \le x+dx, y \le Y \le y+dy\}}{P\{y \le Y \le y+dy\}}$
    $= P\{x \le X \le x + dx | y \le Y \le y + dy\}$
  - In other words, for small values dx and dy, $f_{X|Y}(x|y)\ dx$ is approximately the conditional probability that X is between x and $x + dx$ given that Y is between y and $y + dy$.
- The *conditional expectation* of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by
  - $E[X|Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)dx$

Example 1

- **Example 3.6** Suppose the joint density of X and Y is given by

$$f(x,y) = \begin{cases} 6xy(2-x-y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Compute the conditional expectation of X given that $Y = y$, where $0 < y < 1$.

Example 2

- **Example 3.7** Suppose the joint density of X and Y is given by

$$f(x, y) = \begin{cases} 4y(x-y)e^{-(x+y)}, & 0 < x < \infty, 0 \leqslant y \leqslant x \\ 0, & \text{otherwise} \end{cases}$$

Compute $E[X|Y = y]$.

Example 3

- **Example 3.8** The joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{2} y e^{-xy}, & 0 < x < \infty, 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

What is $E[e^{\frac{X}{2}}|Y=1]$?

Computing Expectations by Conditioning

- Let us denote by $E[X|Y]$ that function of the random variable Y whose value at $Y=y$ is $E[X|Y=y]$.
- An extremely important property of conditional expectation is that for all random variables X and Y

$$E[X]=E[E[X|Y]]$$

  - Discrete RV: $E[X]=E\big[E[X|Y]\big]=\sum_y E[X|Y=y]P\{Y=y\}$
  - Continuous RV: $E[X]=E\big[E[X|Y]\big]=\int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)dy$
  - Proof?

- Compound random variable
  - The random variable $\sum_{i=1}^{N} X_i$ equal to the sum of a random number N of independent and identically distributed random variables that are also independent of N. the expected value of a compound random variable is $E[X]E[N]$. See examples

Example 1

- **Example 3.10** Sam will read either one chapter of his probability book or one chapter of his history book. If the number of misprints in a chapter of his probability book is Poisson distributed with mean 2 and if the number of misprints in his history chapter is Poisson distributed with mean 5, then assuming Sam is equally likely to choose either book, what is the expected number of misprints that Sam will come across?

Example 2

- **Example 3.11 (The Expectation of the Sum of a Random Number of Random Variables)** Suppose that the expected number of accidents per week at an industrial plant is four. Suppose also that the numbers of workers injured in each accident are independent random variables with a common mean of 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of injuries during a week?

Example 3

- **Example 3.12 (The Mean of a Geometric Distribution)** A coin, having probability p of coming up heads, is to be successively flipped until the first head appears. What is the expected number of flips required?

Example 4

- **Example 3.15** Independent trials, each of which is a success with probability p, are performed until there are k consecutive successes. What is the mean number of necessary trials?

Computing Variances by Conditioning

- Conditional expectations can also be used to compute the variance of a random variable. Specifically, we can use
$$Var(X) = E[X^2] - (E[X])^2$$
and then use conditioning to obtain both E[X] and $E[X^2]$.

- **Example 3.18 (Variance of the Geometric Random Variable)** Independent trials, each resulting in a success with probability *p*, are performed in sequence. Let *N* be the trial number of the first success. Find Var*(N)*.

Computing Variances by Conditioning

- Another way to use conditioning to obtain the variance of a random variable is to apply the conditional variance formula. The conditional variance of X given that $Y = y$ is defined by
$$Var(X|Y=y) = E[(X - E[X|Y=y])^2|Y=y]$$

- That is, the conditional variance is defined in exactly the same manner as the ordinary variance with the exception that all probabilities are determined conditional on the event that $Y = y$. Expanding the right side of the preceding and taking expectation term by term yields
$$Var(X|Y=y) = E[X^2|Y=y] - (E[X|Y=y])^2$$

- Letting Var$(X|Y)$ denote that function of Y whose value when $Y = y$ is Var$(X|Y = y)$, we have the following result.

- **Proposition 3.1 (The Conditional Variance Formula)**
$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

Computing Variances by Conditioning

- **Example 3.19 (The Variance of a Compound Random Variable)** Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with distribution F having mean μ and variance σ^2, and assume that they are independent of the nonnegative integer valued random variable N. As noted in Example 3.11, where its expected value was determined, the random variable $S = \sum_{i=1}^{N} X_i$ is called a compound random variable. Find its variance.

Computing Probabilities by Conditioning

- Not only can we obtain expectations by first conditioning on an appropriate random variable, but we may also use this approach to compute probabilities.

- To see this, let E denote an arbitrary event and define the indicator random variable X by

$$X = \begin{cases} 1, \text{if } E \text{ occurs} \\ 0, \text{if } E \text{ does not occur} \end{cases}$$

- It follows from the definition of X that
  $E[X] = P(E)$
  $E[X|Y = y] = P(E|Y = y)$, for any random variable Y

- Therefore, we obtain
  $P(E) = \sum_y P[E|Y = y]P\{Y = y\}$, if Y is discrete
  $= \int_{-\infty}^{\infty} P[E|Y = y]f_Y(y)dy$, if Y is continuous

Examples

- **Example 3.21** Suppose that X and Y are independent continuous random variables having densities f_X and f_Y, respectively. Compute P{X < Y}.

- **Example 3.22** An insurance company supposes that the number of accidents that each of its policyholders will have in a year is Poisson distributed, with the mean of the Poisson depending on the policyholder. If the Poisson mean of a randomly chosen policyholder has a gamma distribution with density function

$$g(\lambda) = \lambda e^{-\lambda},\ \lambda \geq 0$$

  what is the probability that a randomly chosen policyholder has exactly n accidents next year?

Examples

- **Example 3.28** Let $U_1, U_2, \ldots$ be a sequence of independent uniform (0, 1) random variables, and let
$$N = \min\{n \geq 2: U_n > U_{n-1}\} \text{ and}$$
$$M = \min\{n \geq 2: U_1 + \cdots + U_n > 1\}$$
That is, N is the index of the first uniform random variable that is larger than its immediate predecessor, and M is the number of uniform random variables we need sum to exceed 1.

- **Example 3.29** Let $X_1, X_2, \ldots$ be independent continuous random variables with a common distribution function F and density $f = F'$, and suppose that they are to be observed one at a time in sequence. Let
$$N = \min\{n \geq 2: X_n = \text{second largest of } X_1, \ldots, X_n\} \text{ and}$$
$$M = \min\{n \geq 2: X_n = \text{second largest of } X_1, \ldots, X_n\}$$
Which random variable—X_N, the first random variable which when observed is the second largest of those that have been seen, or X_M, the first one that on observation is the second smallest to have been seen—tends to be larger?