

## 7.6 조건부기댓값과 예측 문제

확률변수  $Y$ 의 기댓값을 구할 때 주변 확률밀도함수  $p_Y(y)$ 를 사용하여 가중치를 계산하지 않고 조건부 확률밀도함수  $p_{Y|X}(y|x)$ 를 이용하여 가중치를 계산하면 **조건부기댓값(conditional expectation)** 혹은 **조건부평균(conditional mean)**이 된다.

$$E_Y[Y|X] = \int_{y=-\infty}^{y=\infty} y p_{Y|X}(y|x) dy \quad (7.6.1)$$

또는 간단히 다음처럼 쓴다.

$$E[Y|X] = \int y p(y|x) dy \quad (7.6.2)$$

조건부기댓값에서 조건이 되는 확률변수  $X$ 의 값  $x$ 는 조건부기댓값을 사용하는 사용자가 지정해야 하는 독립변수이다. 따라서 **조건부기댓값은 조건이 되는 확률변수의 값에 따라서 값이 달라지는 확률변수**이다. 또한  $E[Y|X]$ 는 조건이 되는 확률변수  $X$ 의 값  $x$ 를 입력으로 가지는 함수다.

$$E[Y|X = x] = f(x) \quad (7.6.3)$$

이 식은 간단히 다음처럼 쓰기도 한다.

$$E[Y|x] = f(x) \quad (7.6.4)$$

$f(x)$ 는 조건이 되는 확률변수  $X$ 의 값  $x$ 를 입력받아서 결과가 되는 확률변수  $Y$ 의 기댓값을 출력하는 함수다.

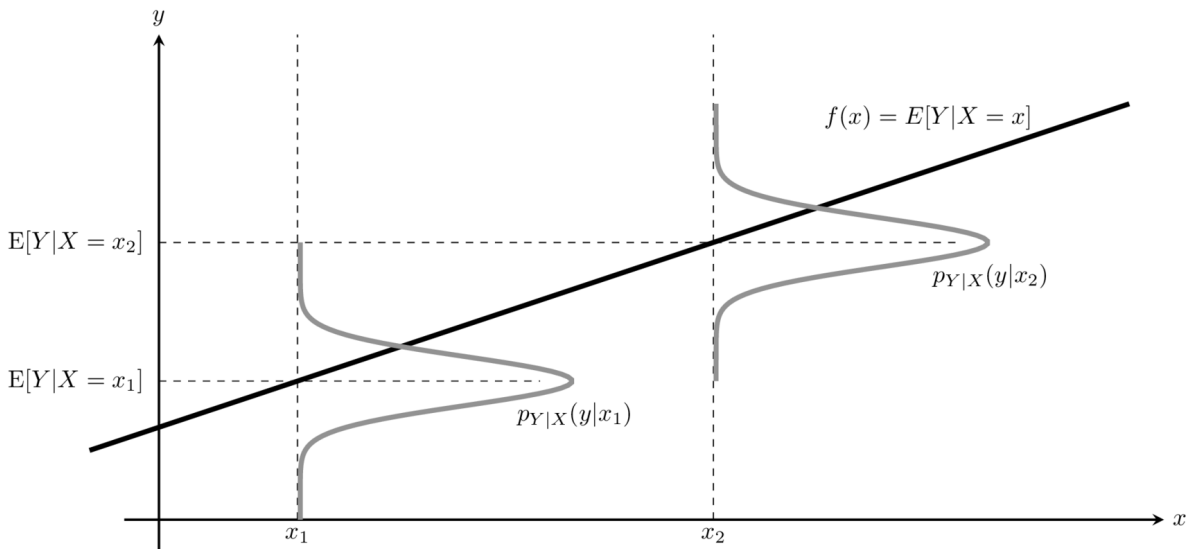


그림 7.6.1 : 조건부기댓값

### 예측 문제

두 확률변수가  $X, Y$ 에서  $X$ 의 값을 알고 있을 때  $Y$ 의 값을 알아내는 것을 **예측(prediction)**문제라고 한다.  $Y$ 가 연속확률변수면 회귀분석(regression analysis),  $Y$ 가 이산확률변수면 분류(classification)라고도 한다.  $X$ 의 값

$x$ 을 알면 조건부확률분포  $P(y|x)$ 의 분포를 알 수 있지만 가장 대표성이 있는 하나의 값이 되어야하므로 일반적으로 조건부확률분포의 기댓값인 조건부기댓값을 예측문제의 답으로 하는 경우가 많다. 경우에 따라서는 예측문제의 답으로 중앙값이나 최빈값 등을 계산할 수도 있다. 예측문제의 답은  $\hat{y}$ 기호로 표기하기도 한다. 조건부기댓값은  $x$ 값의 함수이므로 이 함수를 구할 수 있으면 모든  $x$ 값에 대한 예측결과를 구한 것과 같다.

$$x \xrightarrow{\text{예측}} \hat{y} = E[y|x] = f(x) \quad (7.6.5)$$

## 조건부기댓값의 성질

조건부기댓값  $E[Y|X]$ 가  $X$ 의 함수, 즉 변환(transform)이므로 **조건부기댓값  $E[Y|X]$ 도 확률변수다.**

만약 확률변수  $Y$ 가 확률변수  $X$ 의 값을 독립변수로 하는 결정론적 함수값이라면

$$Y = g(X) \quad (7.6.6)$$

사용자가  $X$ 의 값을 어떤 값  $x$ 로 정하는 순간  $Y$ 의 값도 결정되어 버리기 때문에  $Y = g(X)$ 는 더이상 확률적인 값이 아니라 상수가 된다.

$$E[Y|X] = E[g(X)|X] = g(X) \quad (7.6.7)$$

같은 방식으로 확률변수  $X$ 와  $Y$ 가 결정론적 함수 관계가 아닐 때도 다음 등식이 성립한다.

$$E[g(X)Y|X] = g(X)E[Y|X] \quad (7.6.8)$$

## 전체 기댓값의 법칙

조건부기댓값은 확률변수이므로 조건이 되는 확률변수에 대해 다시 기댓값을 구할 수 있다. 이렇게 반복하여 구한 조건부기댓값의 기댓값은 원래 확률변수의 기댓값과 같다.

$$E_X[E_Y[Y|X]] = E_Y[Y] \quad (7.6.9)$$

간단히 다음처럼 쓰기도 한다.

$$E[E[Y|X]] = E[Y] \quad (7.6.10)$$

이를 **전체 기댓값의 법칙(law of total expectation)** 또는 **반복 기댓값의 법칙(law of iterated expectation)**이라고 한다.

$X, Y$ 가 이산확률변수인 경우에는 다음처럼 증명할 수 있다.

$$\begin{aligned} E_X[E_Y[Y|X]] &= \sum_{x_i \in X} p(x_i) E_Y[Y|X] \\ &= \sum_{x_i \in X} p(x_i) \sum_{y_j \in Y} p(y_j|x_i) y_j \\ &= \sum_{x_i \in X} \sum_{y_j \in Y} p(x_i) p(y_j|x_i) y_j \\ &= \sum_{x_i \in X} \sum_{y_j \in Y} p(x_i, y_j) y_j \\ &= \sum_{y_j \in Y} p(y_j) y_j \\ &= E_Y[Y] \end{aligned} \quad (7.6.11)$$

### 연습 문제 7.6.1

전체 기댓값의 법칙을 사용하여 다음을 증명하라.

$$E[(Y - E[Y|X])g(X)] = 0 \quad (7.6.12)$$

## 조건부분산

조건부기댓값을 정의한 것처럼 조건부분산(conditional variance)도 다음처럼 정의할 수 있다.

$$\text{Var}_Y[Y|X] = E_Y[(Y - E_Y[Y|X])^2|X] = \int (Y - E_Y[Y|X])^2 f_{Y|X}(y|x) dy \quad (7.6.13)$$

조건부분산은  $x$ 의 값을 알고 있을 때 이에 대한 조건부확률분포  $p(y|x)$ 의 분산이다.

예측문제의 관점으로 보면 조건부분산은 예측의 불확실성, 즉 예측으로 맞출 수 없는 범위를 뜻한다.

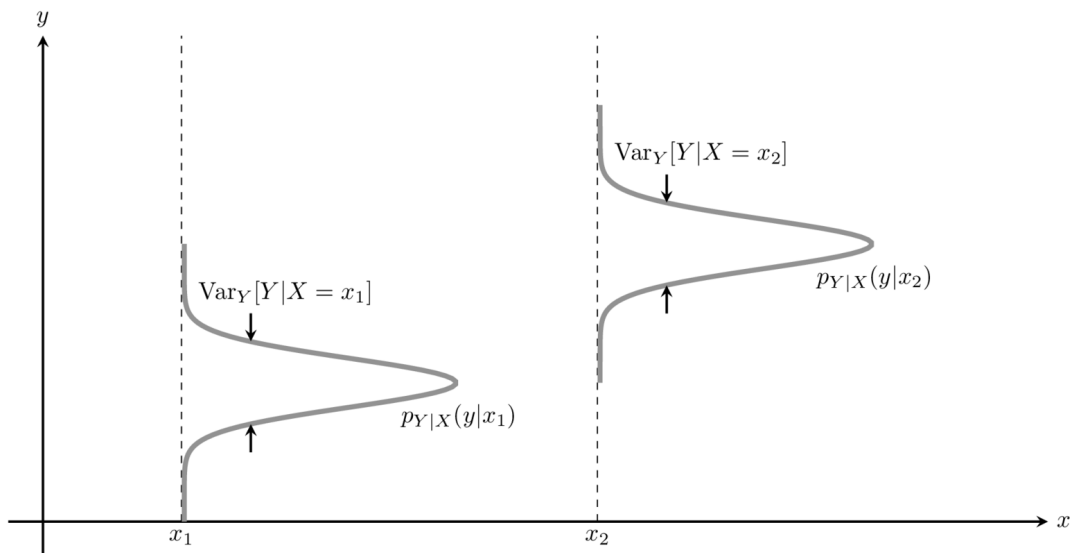


그림 7.6.2 : 조건부분산

## 전체 분산의 법칙

확률변수의 분산은 조건부분산의 기댓값과 조건부기댓값의 분산의 합과 같다. 이를 **전체 분산의 법칙**(law of total variance)라고 한다.

$$\text{Var}[Y] = E[\text{Var}[Y|X]] + \text{Var}[E[Y|X]] \quad (7.6.14)$$

전체 기댓값의 법칙을 사용하여 증명할 수 있다.

$$\begin{aligned}
\text{Var}[Y] &= E[Y^2] - (E[Y])^2 \\
&= E[E[Y^2 | X]] - (E[E[Y | X]])^2 \\
&= E[\text{Var}[Y | X] + (E[Y | X])^2] - (E[E[Y | X]])^2 \\
&= E[\text{Var}[Y | X]] + (E[(E[Y | X])^2] - (E[E[Y | X]])^2) \\
&= E[\text{Var}[Y | X]] + \text{Var}[E[Y | X]]
\end{aligned} \tag{7.6.15}$$

$E[Y | X] = \hat{y}$ 로 표현하면 다음과 같이 쓸 수 있다.

$$\text{Var}[Y] = E[(\hat{y} - y)^2] + \text{Var}[\hat{y}] \tag{7.6.16}$$

예측문제의 관점에서 조건분산의 기댓값  $E[(\hat{y} - y)^2]$ 은 예측 오차 즉, 편향(bias)의 평균적인 크기를 뜻한다. 조건부기댓값의 분산  $\text{Var}[\hat{y}]$ 은 예측값의 변동 크기다. 예측값의 변동 크기가 증가한다는 것은 예측모형이 복잡하고 비선형적이며 주어진 데이터에 과최적화되기 쉽다는 의미다. 따라서 전체 분산의 법칙이 말하고자 하는 바는 예측 오차의 크기와 예측값의 변동의 합이 일정하므로 예측 오차를 줄이면 모형이 복잡해지고 과최적화가 되며 반대로 모형을 과최적화를 막기 위해 단순하게 하면 예측 오차가 증가한다. 이를 **편향-분산 상충(Bias-variance Tradeoff)** 법칙이라고도 한다.