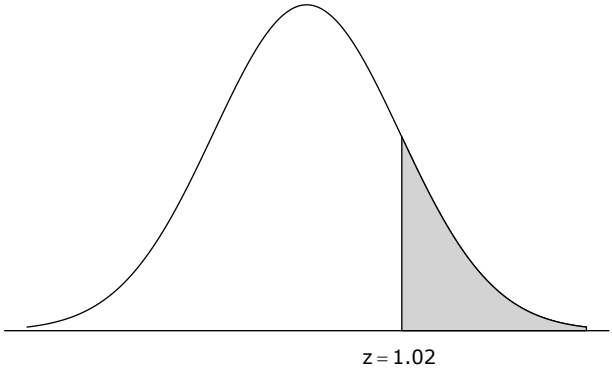


그림 1.5 $z = 1.02$ 보다 오른쪽인 면적

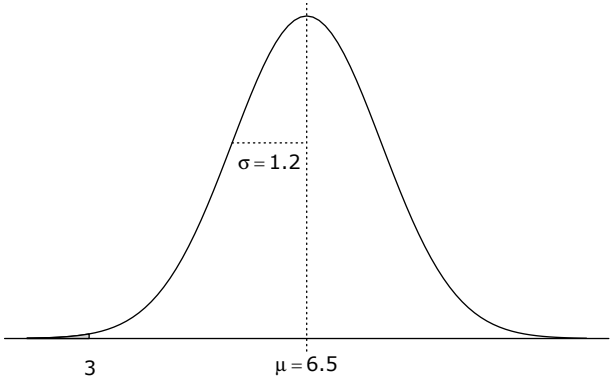


그림 1.6 평균 6.5 표준편차 1.2 인 정규분포

같다.

$$z = \frac{X - \mu}{\sigma} = \frac{3 - 6.5}{1.2} \approx -2.91 \tag{1.23}$$

책 뒤에 제시되어 있는 표준정규분포표에서 $P(z \leq -2.91) = 0.0018$ 이다. ■

이제 백분율에 해당하는 z 값을 찾는 방법에 대해 생각해 보자. 만일 백분위수 P_5 에 해당하는 z 값을 구하려면 표준정규분포표에서 면적이

같이 정리할 수 있다.

1. $n \geq 30$ 일 때 크기 n 의 표본이 평균 μ, 표준편차 σ 를 갖는 모집단으로부터 추출되었다면 표본평균의 표본분포는 근사적으로 정규분포를 따른다.

2. 모집단이 정규분포를 따르면 표본평균의 표본분포는 임의의 표본크기에 대하여 정규분포를 따른다.

즉, 중심극한정리는 모집단이 무슨 분포를 하든 표본 크기 n 이 크면 표본평균 $\bar{x}$ 의 분포는 정규분포를 하며, 표본 크기가 작더라도 모집단이 정규분포를 하면 표본평균의 분포는 정규분포를 한다는 것을 의미한다. 표본 통계량이 표본평균이라고 하면 표 1.9 와 같은 성질이 있다.

표 1.9 표본평균의 성질

1. 표본평균의 평균 $\mu_{\bar{x}}$ 는 모평균 μ 와 같다.

2. 표본평균의 표준편차 $\sigma_{\bar{x}}$ 는 모표준편차를 $\sqrt{n}$ 으로 나눈 것이다.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

예제 1.12 어떤 집단의 평균 키가 172cm, 표준편차가 7cm 이다. 이 모집단에서 42 명의 확률표본을 뽑아 평균 키를 구할 때 표본분포의 평균과 표준오차(standard error) 를 구해 보자.

표본분포의 평균은 모평균과 같기 때문에 $\mu_{\bar{x}} = \mu = 172$ 이다. 표본분포의 표준편차인 평균의 표준오차는 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7}{\sqrt{42}} \approx 1.08$ 이다. 중심극한정리로부터 표본 크기가 30 보다 크므로 표본분포는 $\mu = 172$, $\sigma = 1.08$ 인 정규분포로 근사화할 수 있다. ■

여기서 μ 는 모수로 그 값이 일정하고, 표본 크기 n 은, 표본을 취할 때마다 일정한 크기로 취한다면 이 또한 일정한 값을 갖는다. 따라서 t-분포에서는 표본마다 값이 달라지는 것은 정규분포와 달리 $\bar{x}$ 와 s 가 된다. 그림 2.3 은 표준정규분포와 자유도에 따른 t-분포를 보여 준다.

신뢰도 $1-\alpha$ 인 경우 t-값은 정규분포에서와 마찬가지로 임계값 $-t_{\frac{\alpha}{2}}$ 와 $t_{\frac{\alpha}{2}}$ 사이가 된다. t-분포는 자유도 $n-1$ 에 따라 분포 모양이 결정되므로 $n-1$ 을 붙여, $-t_{\frac{\alpha}{2};n-1}$, $-t_{\frac{\alpha}{2};n-1}$ 로 나타낸다.

표본의 크기가 15 이고 95% 신뢰도 $(c=0.95)$ 의 임계값 $-t_{0.025;14}$, $t_{0.025;14}$ 을 알아보자. 이 책 뒤의 부록으로 제공되는 t-분포표에서 신뢰수준 95% 에서 자유도 14 인 값은 -2.145 와 2.145 이다. 그림 2.4 는 이 분포를 나타낸다.

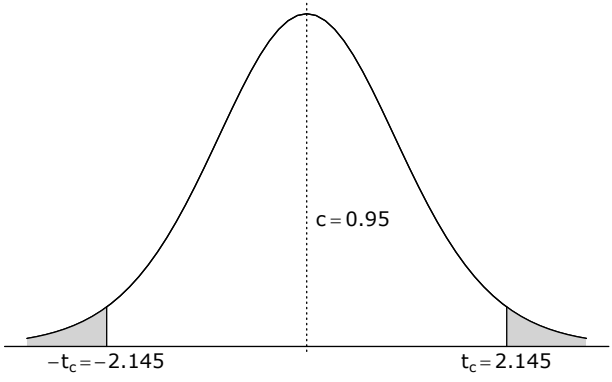


그림 2.4 자유도 14 인 t-분포

2.3 카이제곱 분포

카이제곱(χ^2) 분포 (chi-square distribution) 는 영국의 통계학자 Pearson 이 고안했다. 확률변수의 값을 제곱하기 때문에 음수의 값이 나오지 않는다. 우선 $Z_1, Z_2, ...Z_\phi$ 가 서로 독립인 표준정규분포를 보이는 확률

표 2.2 카이제곱분포의 특성

1. 정규분포처럼 연속형 확률분포다.
2. $Y \sim \chi^2(\phi)$ 일 때, 기댓값은 다음과 같다.

$$E(Y) = E\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{n-1}{\sigma^2}E(s^2) \tag{2.4}$$

   여기서 표본분산 $s^2 = \frac{S}{n-1}$ 의 기댓값은 모분산과 같다. 따라서 $E(s^2) = \sigma^2$ 이고, $n-1 = \phi$ 이다. 즉 표본 크기보다 1 이 적은 자유도 $n-1$ 인 카이제곱분포를 따른다.

3. 양의 왜도 (positive skewness) 를 가진다.

변수일 때 다음과 같이 Y 를 정의한다.

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_\phi^2 \tag{2.5}$$

확률변수 Y 는 자유도 ϕ 인 카이제곱분포라고 하고 다음과 같이 표시한다.

$$Y \sim \chi^2(\phi) \tag{2.6}$$

이제 확률변수 Y 를 계산해 보자. Y 는 (2.5) 에서처럼 표준정규분포의 제곱의 합이므로 다음과 같이 표시할 수 있다.

$$\begin{aligned} Y &= \sum Z_i^2 = \sum \left(\frac{X_i - \mu}{\sigma}\right)^2 \\ &= \frac{1}{\sigma^2}\sum (X_i - \mu)^2 \\ &= \frac{S}{\sigma^2} \end{aligned} \tag{2.7}$$

모분산 σ^2 의 추정량으로 표본분산 $s^2 = \frac{S}{n-1}$ 을 사용하므로 (2.7) 은

2.4.2 신뢰구간을 이용한 가설 검정

귀무가설 또는 대립가설을 세운 후 이 가설을 받아들일지 아니면 기각할지를 결정해야 한다. 일반적으로 표본의 결과로 얻어진 $\bar{x}$ 의 값과 모평균 가정치 μ 와의 차이를 가지고 가설을 기각할지 받아들일지를 결정하게 된다. 이런 경우 판단의 기준이 되는 평가 기준이 필요하다. 표본의 결과로 나온 $\bar{x}$ 의 값이 어느 정도일 때 귀무가설을 채택할 수 있는지 판가름할 수 있는 평가 기준이 필요하다. 이 평가 기준에는 크게 신뢰구간을 사용하는 것과 검정통계량(test statistic)을 이용하는 방법이 있다. 우선 신뢰구간을 이용하는 경우를 살펴보자.

예제 2.5 어느 인삼 드링크에 사포닌 함량이 20mg 이라고 표기되어 있는데, 이 함량이 드링크마다 적절한지, 즉 많지도 않고 적지도 않은지를 알아보기 위하여 인삼 드링크 100 개를 표본으로 하여 표본평균을 구해보니 $\bar{x}$=19.5mg 이 나왔다. 일반적으로 모표준편차인 σ 는 2.0mg 으로 알려져 있다. $\alpha = 0.05$ 로 하여 검정해 보자.

이 문제에서 사포닌 함량이 20mg 이라면 아무 문제가 되지 않기 때문에, 연구자가 밝히고 싶은 것은 의약품의 함량이 20mg 이 아닐 수도 있다는 것이다. 따라서 귀무가설은 $H_0 : \mu = 20$ 이고 대립가설은 $H_a : \mu \neq 20$ 이다. 신뢰구간을 이용한 가설 검정은 모평균 가정치 μ_0 가 신뢰구간 내에 들어오지 않으면 귀무가설을 기각한다. 따라서 다음과 같은 신뢰하한치와 신뢰상한치를 임계치로 사용한다.

$$\bar{x} - z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \leq \mu_0 \leq \bar{x} + z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \tag{2.9}$$

여기서 $\bar{x} = 19.5$, $n = 100$, $\sigma = 2.0$ 이고, $z_{0.025} = 1.96$, $\sigma_{\bar{x}} = \frac{2}{\sqrt{100}} = 0.2$ 이다. 따라서

$$x \pm z_{\alpha/2}\sigma_{\bar{x}} = 19.5 \pm 1.96 \times 0.2 = 19.5 \pm 0.392$$

$$19.108 \leq \mu \leq 19.892$$

표 2.4 대립가설과 검정 방법

1. 대립가설 H_a 이 부등호 $<$ 를 포함하면 좌측검정(left-tailed test) 을 행한다.
2. 대립가설 H_a 이 부등호 $>$ 를 포함하면 우측검정(right-tailed test) 을 행한다.
3. 대립가설 H_a 이 $\neq$ 를 포함하면 양측검정(two-tailed test) 을 행한다.

예제 2.8 어떤 공장에서 가동하는 기계 부품의 평균 수명은 30 년보다 길다고 주장한다. 임의로 추출한 부품 36 개는 표본평균 31.5 년이고 표준편차 3.5 년이다. 유의수준 $\alpha = 0.01$ 에서 위의 주장을 충분히 뒷받침하는 충분한 근거가 있는지 알아보자.

기계 부품의 평균 수명을 μ 로 하면, 다음과 같은 가설이 성립한다.

$$H_0 : \mu \leq 30$$
$$H_a : \mu > 30$$

z_0 검정통계량을 구하면 다음과 같다.[3]

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{31.5 - 30}{3.5/\sqrt{36}} = 2.57$$

대립가설이 $>$ 를 포함하기 때문에 우측검정에 해당한다. p-값은 $z_0 = 2.57$ 의 오른쪽 영역이다. 따라서 p-값은 $0.0051(1 - 0.9949)$ 이다. 이 값은 $\alpha = 0.01$ 보다 작기 때문에 귀무가설을 기각한다. 따라서 유의수준 1% 에서 부품의 평균수명은 30 년보다 길다는 주장은 충분히 근거가 있다. ■

[3] $n \geq 30$ 이면 $\sigma \approx s = 3.5$ 이다.

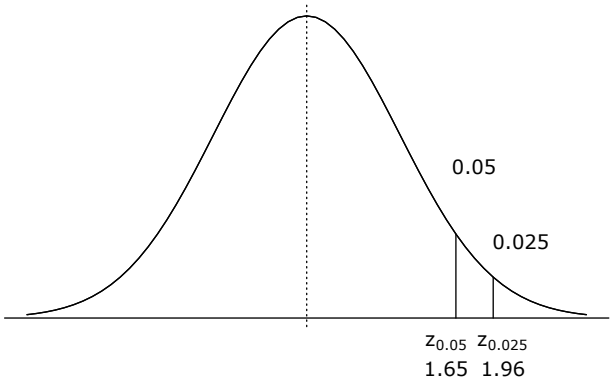


그림 2.11 양측검증과 단측검증 ($\alpha = 0.05$)

예제 2.9 새로운 체중조절 프로그램에 대한 광고의 참가자는 평균 12주 이내에 10kg을 감량할 수 있다고 한다. 참가자 중 임의로 선택한 60명의 한 달 동안의 감량을 조사한 결과 10kg을 감량하는 데 평균 11.2주, 표준편차 3.4주로 나타났다. 유의수준 $\alpha = 0.05$에서 이 주장을 뒷받침하는 충분한 근거가 있는지 살펴보자.

참가자의 평균 감량 시간을 μ로 하면, 다음과 같은 가설이 성립한다.

$$H_0 : \mu = 12$$
$$H_a : \mu < 12$$

z_0 검정통계량을 구하면 다음과 같다.

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{11.2 - 12}{3.4/\sqrt{60}} = -1.82$$

이 검정은 좌측검정에 해당한다. p-값은 $z_0 = -1.82$의 왼쪽 영역이고 그 값은 0.0344이다. 이 값은 $\alpha = 0.05$보다 작기 때문에 귀무가설을 기각한다. ■

바탕으로 한 각 단어의 확률은 다음과 같다.

$$P(\text{new}) = \frac{15828}{14307668}$$
$$P(\text{companies}) = \frac{4675}{14307668}$$

이제 'new'와 'companies'가 서로 관련이 없이 독립적으로 출현한다고, 즉 연어 구성이 아니라고 귀무가설을 설정한다. 독립적으로 출현하기 때문에 다음과 같은 확률이 설정된다.

$$\begin{aligned} H_0 : P(\text{new companies}) &= P(\text{new})P(\text{companies}) \\ &= \frac{15828}{14307668} \times \frac{4675}{14307668} \\ &\approx 3.615 \times 10^{-7} \end{aligned}$$

귀무가설이 참이라면 임의적으로 구성되는 바이그램에서 'new companies'가 나오면 성공, 그렇지 않으면 실패로 하는 베르누이 시행이라고 생각할 수 있다. 이 시행에서 성공할 확률은 3.615×10^{-7} 이다. 이는 평균은 $\mu = 3.615 \times 10^{-7}$ 이고 분산은 $\sigma^2 = p(1-p)$ 인 이항분포다.[5] 실제로 이 바이그램의 확률은 아주 작기 때문에 $\sigma^2 = p(1-p) \approx p$ 로 근사화한다.

이제 t-값을 계산해 보자. 'new companies'는 14,307,668 바이그램에서 8번 나타난다. 따라서 평균은 $\bar{x} = \frac{8}{14,307,668} \approx 5.591 \times 10^{-7}$ 이다. 이를 바탕으로 t-값을 구하면 다음과 같다.

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{N}}} \approx \frac{5.591 \times 10^{-7} - 3.615 \times 10^{-7}}{\sqrt{\frac{5.591 \times 10^{-7}}{14307668}}} \approx 0.999$$

$\alpha = 0.005$ 이고 자유도 ∞ 인 t-값은 2.576 이다.[6] 0.999는 기각역

[5] 이항분포의 평균과 분산은 각각 np, $np(1-p)$ 이다. 여기서는 n 을 1로 하여 계산하였다.

[6] t-분포표는 대개 자유도가 1에서 30까지인 경우와 그 다음은 ∞ 로 값이 제시된다.

예제 2.16 어느 제강회사에서 동선코일의 인장강도를 테스트하기 위하여 A 코일 14 개와 B 코일 12 개를 조사하였더니 각각 평균 120 과 표준편차 8, 평균 116 과 표준편차 10 이 나왔다. 동선코일의 인장강도는 정규분포를 따른다고 할 때 코일 A 가 더 인장강도가 높다고 할 수 있는지를 99% 신뢰구간에서 검정해 보자.

우선 A, B 의 평균 인장강도를 각각 μ_1, μ_2 라고 하면 다음과 같은 귀무가설을 설정할 수 있다.

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 > 0$$

(2.18) 에 의해 통합표준편차 (s_p) 를 다음과 같이 구할 수 있다.

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$
$$= \sqrt{\frac{(14-1)8^2 + (12-1)10^2}{14+12-2}} \approx 8.97$$

표준편차는 다음과 같다.

$$\sigma_{\bar{x}_1-\bar{x}_2} = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 8.97\sqrt{\frac{1}{14}+\frac{1}{12}} \approx 3.52$$

이를 바탕으로 t-값은 다음과 같이 구해진다.

$$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} = \frac{120-116}{3.52} \approx 1.13$$

자유도는 $24(=14-1+12-1)$ 이고, $\alpha = 0.05$ 이다. 이에 해당하는 t-값은 1.711 이다. 이 검정은 우측검정이기 때문에 기각역은 $t > 1.711$ 이다. 구해진 t-검정량 1.13 은 기각역에 있지 않기 때문에 귀무가설을 받아들여야 한다. 따라서 코일 A 가 인장강도가 더 높다는 주장은 충분한 근거가 없다. ■

이 예의 전체 제곱합은 다음과 같이 계산된다.

$$\begin{aligned} SST &= (91-72.4)^2 + (72-72.4)^2 + (85-72.4)^+(81-72.4)^2 \\ &\quad + (66-72.4)^2 + (80-72.4)^2 + (72-72.4)^2 + (65-72.4)^2 \\ &\quad + (78-72.4)^2 + (69-72.4)^2 + (71-72.4)^2 + (64-72.4)^2 \\ &\quad + (58-72.4)^2 + (70-72.4)^2 + (64-72.4)^2 \\ &= 1131.6 \end{aligned}$$

여기서 전체 제곱합은 총 15 개의 평균치에 대한 15 개의 점수들의 편차에 기초하기 때문에 $14(=15-1)$ 의 자유도를 갖는다. 이 자유도는 다음과 같이 정의된다.

$$\text{전체 자유도}: \qquad df_t = n_{\text{전체}} - 1 \tag{3.6}$$

자세히 살펴보면 전체 제곱합(SST)은 집단내 제곱합(SSW)과 집단간 제곱합(SSB)의 합임을 알 수 있다. 따라서 앞에서 계산된 전체 제곱합 1131.6은 $SSW + SSB$ 의 합 $668 + 463.6$ 과 같다. 이를 정리하면 다음과 같다.

$$\begin{aligned} SST &= \sum (X - \bar{\bar{X}})^2 = \sum (X - \bar{X} + \bar{X} - \bar{\bar{X}})^2 \\ &= \sum \left\{ (X-\bar{X})^2 + (\bar{X}-\bar{\bar{X}})^2 + 2(X-\bar{X})(\bar{X}-\bar{\bar{X}}) \right\} \\ &= \sum (X-\bar{X})^2 + \sum (\bar{X}-\bar{\bar{X}})^2 + 2(\bar{X}-\bar{\bar{X}}) \sum (X-\bar{X}) \\ &= \sum (X-\bar{X})^2 + \sum (\bar{X}-\bar{\bar{X}})^2, \qquad \because \sum (X-\bar{X}) = 0 \\ &= SSW + SSB \end{aligned}$$

이제 집단내 제곱합 SSW와 집단간 제곱합 SSB를 각각의 자유도로 나누면 각각 집단내 분산 추정치 s_W^2 와 집단간 분산 추정치 s_B^2 를 구할

표 3.4 일원분산분석 정리

요인	제곱합	자유도	분산추정치	F-비
집단간	$\sum_{\text{모든 점수들}} (\bar{X} - \bar{\bar{X}})^2$	$k-1$	$s_b^2 = \frac{SSB}{df_b}$	$\frac{s_b^2}{s_W^2}$
집단내	$\sum_{\text{모든 점수들}} (X - \bar{X})^2$	$n_{\text{전체}} - k$	$s_W^2 = \frac{SSW}{df_w}$	
전체	$\sum_{\text{모든 점수들}} (X - \bar{\bar{X}})^2$	$n_{\text{전체}} - 1$		

이제 모집단들 차이에 대한 신뢰구간을 어떻게 설정할 수 있는지를 살펴보자. 각 모집단 사이의 차에 관한 신뢰구간은 다음과 같이 표본평균차와 정직한 유의차로 구할 수 있다.

$$|\bar{X}_i - \bar{X}_j| \pm HSD \tag{3.9}$$

여기서의 예, $\mu_1 - \mu_2$ 의 신뢰구간을 구해 보자.

$$|\bar{X}_1 - \bar{X}_2| \pm HSD = |79 - 72.8| \pm 12.57 = 6.2 \pm 12.57$$

따라서 1 학년과 2 학년의 평균 차이는 -6.37 에서 18.77 사이 어디에 있으리라고 95% 확신할 수 있다. 지금까지 살펴본 일원분산분석을 정리하면 표 3.4 와 같다.

3.1.2 이원분산분석

일원분산분석은 하나의 요인(factor) 또는 독립변수들의 서로 다른 수준(level)을 다루지만 경우에 따라 두 가지 이상의 요인들을 동시에 연구할 필요가 있다. 예를 들어 어떤 의사가 우울증을 치료할 때 사용하는 두 치료법의 상대적인 효과와, 그것이 남녀 성별과 관련되는지를 연구하고자

상호작용의 제곱합($SS_{A\times B}$)은 상호작용이 없다고 기대할 때의 칸 평균값들로부터 실제로 얻어진 각각의 평균값들의 제곱으로 구할 수 있다. 이는 지금까지 구한 값들에서 쉽게 계산할 수 있다.

$$SS_{A\times B} = SS_T - (SS_W + SS_A + SS_B) \tag{3.22}$$

이에 해당하는 자유도는 df_A 와 df_B 를 곱한 값이다.

$$df_{A\times B} = (R-1)(C-1) \tag{3.23}$$

이 예의 $df_{A\times B} = 1 = (2-1)(2-1)$ 이다.

이제 F-검정을 위해 분산추정, 즉 모분산 분석을 행한다. 이를 위해 (3.12)에서 살펴보았듯이 네 가지 제곱합들 — SS_W, SS_A, SS_B, $SS_{A\times B}$—을 각각 자신의 자유도로 나누어 각각의 분산을 추정한다. 이 예의 분산 추정치는 다음과 같다.

$$s_W^2 = \frac{SSW}{df_w} = \frac{70}{12} \approx 5.83$$

$$s_A^2 = \frac{SS_A}{df_A} = \frac{36}{1} = 36$$

$$s_B^2 = \frac{SS_B}{df_B} = \frac{4}{1} = 4$$

$$s_{A\times B}^2 = \frac{SS_{A\times B}}{df_{A\times B}} = \frac{1}{1} = 1$$

이제 s_A^2, s_B^2, $s_{A\times B}^2$ 을 집단 내 s_W^2 으로 나누어 F-비를 구할 수 있다.

$$F_A = \frac{s_A^2}{s_W^2} = \frac{36}{5.83} \approx 6.17$$

$$F_B = \frac{s_B^2}{s_W^2} = \frac{4}{5.83} \approx 0.68$$

표 3.8 이원분산분석 정리

분산요인	제곱합	자유도	분산추정치	F-비
집단간	$SS_B = SS_A + SS_B + SS_{A\times B}$			
A	$\frac{\left(\sum X_{A_1}\right)^2 + \left(\sum X_{A_2}\right)^2 + \cdots}{n_A} - \frac{\left[\sum_{\text{모든 점수들}} X\right]^2}{n_{\text{전체}}}$	$R-1$	$\frac{SS_A}{df_A}$	$\frac{s_A^2}{s_W^2}$
B	$\frac{\left(\sum X_{B_1}\right)^2 + \left(\sum X_{B_2}\right)^2 + \cdots}{n_B} - \frac{\left[\sum_{\text{모든 점수들}} X\right]^2}{n_{\text{전체}}}$	$C-1$	$\frac{SS_B}{df_B}$	$\frac{s_B^2}{s_W^2}$
$A\times B$	$SS_{A\times B} = SS_T - (SS_W + SS_A + SS_B)$	$(R-1)(C-1)$	$\frac{SS_{A\times B}}{df_{A\times B}}$	$\frac{s_{A\times B}^2}{s_W^2}$
집단내	$\sum_{\text{모든 점수들}} X^2 - \frac{\sum_{\text{모든 칸들}} \left[\sum_{\text{칸}} X\right]^2}{n_{\text{칸}}}$	$RC(n_{\text{칸}}-1)$	$\frac{SSW}{df_w}$	
전체	$\sum_{\text{모든 점수들}} X^2 - \frac{(\sum X)^2}{n_{\text{전체}}}$	$n_{\text{전체}}-1$		

표 3.9 반복측정 일원분산분석

자료	제곱합	자유도	분산추정치	F-비
개체간(S)	$\sum k(\bar{X}_{\text{subj}} - \bar{\bar{X}})^2$	$n_{\text{개체}} - 1$	$\frac{SSS}{df_s}$	
집단간(B)	$\sum_{\text{모든 점수들}} (\bar{X} - \bar{\bar{X}})^2$	$k - 1$	$\frac{SSB}{df_b}$	$\frac{s_B^2}{s_r^2}$
잔차(R)	$SST - SSB - SSS$	$df_s \times df_b$	$\frac{SSR}{df_r}$	
전체(T)	$\sum_{\text{모든 점수들}} (X - \bar{\bar{X}})^2$	$n_{\text{전체}} - 1$		

다음의 실험을 하였다.[5] 한국인들이 부정관사 'a' 대신에 'the'를 쓰는 경우가 많은지를 알아보기 위해 20명의 실험자를 대상으로 부정관사 'a'가 나타나야 하는 곳에 정관사 'the'가 나타나는 경우를 부분성(partitive)[6]과 관련하여 세 수준으로 구분하여 실험을 하였다. 즉 명시적인 부분성(explicit partitive)인 경우(1a), 내재적인 부분성(implicit partitive)인 경우(1b), 그리고 부분성이 아닌(non-partitive) 경우(1c)에, 부정관사 'a' 대신 정관사 'the'를 남용하여 쓰는 횟수를 측정하여 'the'의 출현이 부분성과 관련이 있는지를 검정하려고 한다. 귀무가설은 'the'의 출현은 부분성과 관련이 없다는 것으로 다음과 같이 설정된다.

$$H_0 : \mu_{1a} = \mu_{1b} = \mu_{1c}$$
$$H_1 : H_0\text{가 아님}$$

이 실험에 대한 결과는 표 3.10과 같으며 계산 결과는 표 3.11로 정리할 수 있다.

[5] 이 예는 Ko et al.(2006)의 원자료를 구하여 분석하였다.

[6] 부분성(partitive)은 언급되는 대상이 이전 대화에서 도입된 한 집합의 구성원 중의 하나로 정의된다.

표 3.11 관사 사용 오류에 관한 반복측정 분산 계산 결과 표

구분	제곱합	자유도	평균합	F-비
개체간	20.26	19	1.06	
집단간	12.034	2	6.017	7.488
잔차	16.63	38	0.80	
총합	48.93	59		

실험 결과에 의한 F-비는 7.488 이다. 그리고 $\alpha = 0.05$ 에서 $F(2, 38)$ 는 3.25 이다. 실험 결과로 나온 F-비는 기각역에 속하기 때문에 귀무가설은 기각된다. 즉 부정관사 'a' 가 사용되어야 할 곳에 'the' 를 쓰는 경우 부분성에는 차이가 있다고 할 수 있다. 이제 어디서 차이가 나는지를 알기 위해서는 일원분석에서와 마찬가지로 Tukey 의 정직한 유의차 검정을 할 필요가 있다. 일원분산분석과 달리 집단내 평균 제곱합 s_W^2 대신에 s_r^2 을 사용한다.

$$HSD = q\sqrt{\frac{s_r^2}{n}} \tag{3.27}$$

여기서 $s_r^2 \approx 0.80$ 이고 $n = 20$ 이다. q 값을 위한 $\alpha = 0.05$ 에서 $df_r = 38$ 이고 $k = 3$ 인 스튜던트화 값은 3.44 이다.[7] 따라서 정직한 유의차는 다음과 같이 계산된다.

$$HSD = 3.44\sqrt{\frac{0.80}{20}} \approx 0.69$$

귀무가설을 기각하기 위해서는 조건의 평균 차가 0.69 이상은 되어야 한다. 1a 와 1c 의 차이 0.8 과 1b 와 1c 의 차이 1.05 가 해당한다. 명시적으로 부분성인 경우와 내재적으로 부분성인 경우가 부분성이 아닌 경우와 차이가 난다는 것을 보여 준다. 따라서 'the' 의 남용은 부분성과 관련이

[7] 통계책에 $df_r = 38$ 에 대한 정확한 값이 제시되지 않은 경우가 있다. 이 경우 $df_r = 40$ 을 사용하였다. 그 직전의 값, $df_r = 30$ 은 3.49 이다.

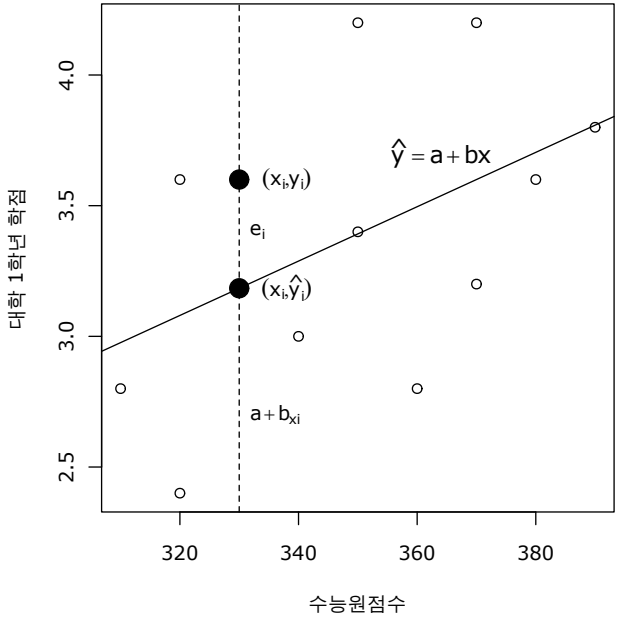


그림 3.5 수능점수와 학점의 관계

앞서 살펴본 대로 회귀선이 적절하기 위해서는 실제값과 추정값인 직선 상의 $\hat{y}$ 값 차이를 나타내는 잔차(residual error) e_i 가 가장 작을 때의 직선을 구해야 한다. 잔차는 다음과 같이 구해진다.

$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - (a + bx_i) \end{aligned} \tag{3.30}$$

관측치가 n 개일 때 이를 모두 반영하기 위해 잔차의 합을 구해서 최소가 되는 값을 구해야 한다. 그러나 잔차가 $+$, $-$ 로 나타나 서로 상쇄되어 그 합은 0 이 되어 버린다. 이를 해소하기 위해 잔차의 제곱합을 이용한다.

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2 \tag{3.31}$$

음과 같이 최소한 세 비트가 필요하다는 것을 의미한다. ■

말1	말2	말3	말4	말5	말6	말7	말8
001	010	011	100	101	110	111	000

만일 이 확률변수에 대해서 우리가 좀 더 많은 정보를 갖고 있다면 그 불확실성(엔트로피)은 줄어들 것이다. 이제 표 5.1과 같은 8마리의 말의 우승할 확률이 주어졌다고 하자.

표 5.1 8마리 말의 우승 확률

말1	말2	말3	말4	말5	말6	말7	말8
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$

표 5.1의 엔트로피는 다음과 같다.

$$\begin{aligned} H(X) &= -\sum_{i=1}^{i=8} P(i)\log P(i) \\ &= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{16}\log\frac{1}{16} - 4(\frac{1}{64}\log\frac{1}{64}) \\ &= 2\text{bits} \end{aligned}$$

개별 말의 우승 확률의 경우, 더 많은 정보가 주어지는 경우에 그 불확실성은 낮아짐을 알 수 있다. 이제 우승 확률이 높은 말은 더 적은 수의 비트로 낮은 말은 더 많은 수의 비트로 전송하여 보낼 수 있다. 가장 확률이 높은 말은 가장 짧은 비트 0으로, 다음은 10, 그 다음은 점점 더 긴 비트로 하여, 110, 1110, 111100, 111101, 111110, 111111로 전송하면, 평균 2비트가 필요함을 알 수 있다. (5.1)의 엔트로피 공식에서 음수

이 주변확률은 음절 단위로 되어 있다. 따라서 자소별 확률은 이 음절 기반의 주변 확률에 $\frac{1}{2}$ 을 곱하여 구할 수 있다. 즉 자음 'p' 의 주변확률 $\frac{1}{8}$, 't' 의 주변확률 $\frac{3}{4}$, 'k' 의 주변확률 $\frac{1}{8}$, 'a' 의 주변확률 $\frac{1}{2}$, 'i' 의 주변확률 $\frac{1}{4}$, 'u' 의 주변확률 $\frac{1}{4}$ 은 음절 단위의 관점에서 보면 중복되어 있기 때문에, $\frac{1}{2}$ 을 곱하여 표 5.4 와 같이 구할 수 있다.

표 5.4 음절 단위에서의 글자 확률

글자	p	t	k	a	i	u
확률	$\frac{1}{16}$	$\frac{3}{8}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

이제 음절을 고려한 자음과 모음의 결합 엔트로피 $H(C,V)$ 를 구해 보자. (5.10) 을 적용하기 위해서는 $H(C)$ 와 $H(V|C)$ 도 필요하다.

$$
\begin{aligned}
H(C) &= 2 \times \frac{1}{8} \times 3 + \frac{3}{4}(2 - \log_2 3) \\
&= \frac{9}{4} - \frac{3}{4}\log 3 \text{bits} \\
&\approx 1.061 \text{bits}
\end{aligned}
$$

$$
\begin{aligned}
H(V|C) &= \sum_{c=p,t,k} p(C=c) H(V|C=c) \\
&= \frac{1}{8} H\left(\frac{1}{2}, \frac{1}{2}, 0\right) + \frac{3}{4} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{8} H\left(\frac{1}{2}, 0, \frac{1}{2}\right) \\
&= 2 \times \frac{1}{8} \times 1 + \frac{3}{4}\left(\frac{1}{2} \times 1 + 2 \times \frac{1}{4} \times 2\right) \\
&= \frac{1}{4} + \frac{3}{4} \times \frac{3}{2} \\
&= \frac{11}{8} \text{bits} \\
&= 1.375 \text{bits}
\end{aligned}
$$

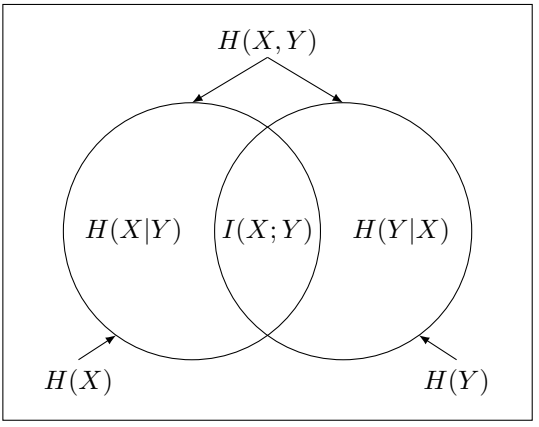


그림 5.1 상호 정보와 엔트로피의 관계

일반적으로 상호 정보 $I(X;Y)$ 는 앞에서 살펴본 엔트로피 도출 과정과 관련하여 다음과 같이 규정할 수 있다.

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) \\ &= \sum_x P(x) \log \frac{1}{P(x)} + \sum_y P(y) \log \frac{1}{P(y)} \\ &\quad - \sum_{x,y} P(x,y) \log \frac{1}{P(x,y)} \\ &= \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \end{aligned} \tag{5.14}$$

여기서 $H(X|X) = 0$ 이므로 $H(X) = H(X) - H(X|X) = I(X;X)$ 가 된다. 따라서 엔트로피가 자신의 정보 (self-information) 를 나타내는 기제임을 알 수 있다.

언어 처리에서 많이 쓰이는 상호 정보는 엄밀히 말해서 점수렴 상호 정보 (pointwise mutual information) 다. 즉, 원래의 상호 정보가 두 확률변수 X 와 Y 사이의 정보에 관한 것이라면 점수렴 상호 정보는 두

비율은 통계학에서 승산(odds)이라 불리는 것으로 $\frac{p}{1-p}$로 구해진다. 이 승산비는 0보다 크고 무한대보다 작은 값으로 나타나며, 확률값이 0에 가까우면 작은 값으로, 1에 가까우면 큰 값으로 나타난다. 예를 들어 어떤 사건이 일어날 확률이 0.8이고 일어나지 않을 확률이 0.2라면 일어날 사건의 승산비(odds ratio)는 $\frac{0.8}{0.2} = 4$이다. 이제 이 선형 모형에서 결과 y가 참일 승산은 다음과 같이 구해진다.

$$\frac{P(y = \text{true}|x)}{1 - P(y = \text{true}|x)} = w \cdot f \tag{5.38}$$

승산비는 0과 무한대 사이의 값으로 나타나기 때문에 이 수식의 좌변과 우변은 같지 않다. 즉, 좌변은 0과 무한대 우변은 $-\infty$와 ∞ 사이의 값으로 나타나기 때문에 좌변에 자연로그를 붙여 양쪽이 다 $-\infty$와 ∞ 사이의 값을 취하도록 해야 한다.

$$\ln\left(\frac{P(y = \text{true}|x)}{1 - P(y = \text{true}|x)}\right) = w \cdot f \tag{5.39}$$

승산의 로그를 취한 것을 로짓 함수(logit function)라 한다.

$$\text{logit}(P(x)) = \ln\left(\frac{P(x)}{1 - P(x)}\right) \tag{5.40}$$

$P(y = \text{true})$를 구하기 위해 수식 (5.39)를 전개해 보자.

$$\begin{aligned}
\ln\left(\frac{P(y = \text{true}|x)}{1 - P(y = \text{true}|x)}\right) &= w \cdot f \\
\frac{P(y = \text{true}|x)}{1 - P(y = \text{true}|x)} &= e^{w \cdot f} \\
P(y = \text{true}|x) &= (1 - P(y = \text{true}|x))e^{w \cdot f} \\
P(y = \text{true}|x) &= e^{w \cdot f} - P(y = \text{true}|x)e^{w \cdot f} \\
P(y = \text{true}|x) + P(y = \text{true}|x)e^{w \cdot f} &= e^{w \cdot f} \\
P(y = \text{true}|x)(1 + e^{w \cdot f}) &= e^{w \cdot f}
\end{aligned}$$

필요가 있다. 다음은 이를 위해 설정된 몇 가지 자질이다.[3]

$$f_1(c,x) = \begin{cases} 1, & \text{해당 형태소가 '나' 이고, } c\text{=VX;} \\ 0, & \text{그렇지 않으면.} \end{cases}$$

$$f_2(c,x) = \begin{cases} 1, & \text{이전 형태소 태그가 'EC' 이고, } c\text{=VX;} \\ 0, & \text{그렇지 않으면.} \end{cases}$$

$$f_3(c,x) = \begin{cases} 1, & \text{다음 형태소가 'ETM' 이고, c=VX;} \\ 0, & \text{그렇지 않으면.} \end{cases}$$

$$f_4(c,x) = \begin{cases} 1, & \text{해당 형태소가 '나' 이고, } c\text{=NP;} \\ 0, & \text{그렇지 않으면.} \end{cases}$$

$$f_5(c,x) = \begin{cases} 1, & \text{이전 형태소 태그가 'ETM' 이고, } c\text{=NP;} \\ 0, & \text{그렇지 않으면.} \end{cases}$$

$$f_6(c,x) = \begin{cases} 1, & \text{다음 형태소 태그가 'JX' 이고, } c\text{=NP;} \\ 0, & \text{그렇지 않으면.} \end{cases}$$

여기서의 자질은 예문과 실제 코퍼스에서 좌우에 나타나는 형태소 위주로 설정되었다. 실제로는 자료에 따라 다른 종류의 다양한 자질이 설정될 수 있다. f_1 에서 f_3 까지는 VX 태그를 위한 자질이며, f_5 에서 f_6 은 NP 태그를 위한 자질이다. 자질은 실제 관찰된 자료를 반영해야 하기 때문에 각각의 형태소와 해당 태그를 연결할 수 있는 자질을 예로 들었다. 또 f_5 에서는 이 예문의 '난'이 동사의 관형형으로 쓰이는 것과 대조적으로 대명사 '나'를 자질로 하기 위해 앞에 또 다른 관형형 '던' 구성이 오는 것을 가정하였다. 즉, 관형형이 연달아 나오는 것은 불가능하다고 보고 그럴 경우는 대명사 '나'로 쓰인다는 것을 자질화하였다.

[3] 태그는 세종 코퍼스에서 사용되는 태그다. VX는 동사를, EC는 연결어미를, ETM은 관형형 어미를, NP는 대명사를, JX는 보조조사를 나타낸다.

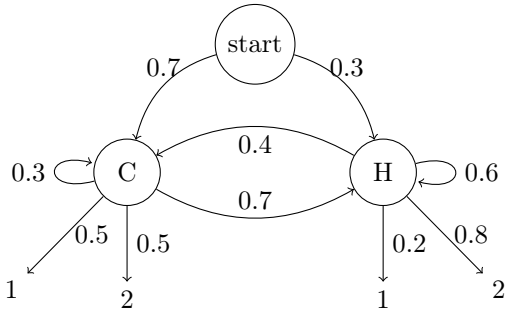


그림 6.5 날씨 상태와 아이스크림 수의 연쇄를 위한 은닉 마르코프 모델

표 6.4 전이 확률과 방출 확률

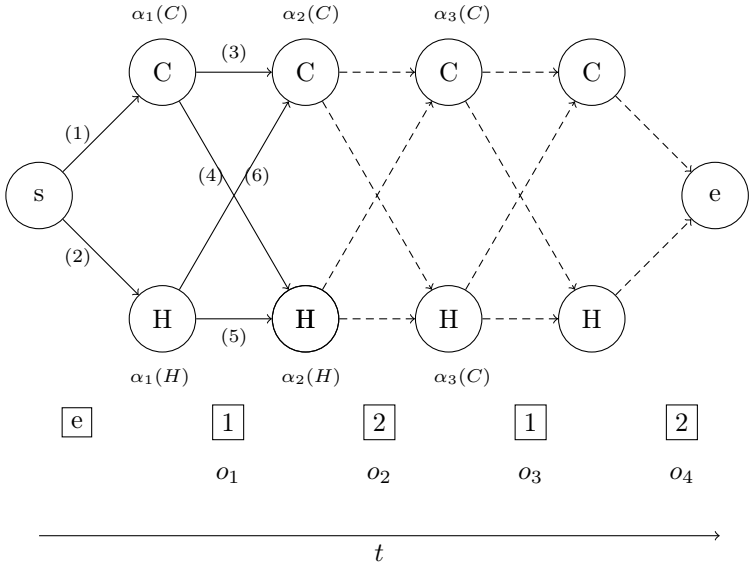
	$P(..\|C)$	$P(..\|H)$	$P(..\|start)$
$P(1\|..)$	0.5	0.2	0.0
$P(2\|..)$	0.5	0.8	0.0
$P(e\|..)$	0.0	0.0	1.0
$P(C\|..)$	0.3	0.4	0.7
$P(H\|..)$	0.7	0.6	0.3

작상태에서 빈글자(empty output)를 출력할 확률을 나타낸다. 이렇게 은닉 마르코프 모델을 설정해 놓고 나면 다음의 세 가지에 대한 문제를 제기할 수 있다.

문제 1(확률의 계산) : 설정된 은닉 마르코프 모델 $\lambda = (A, B)$[1]과 관찰된 연쇄 O가 있을 때 이 모델에서 이 연쇄의 확률 $P(O|\lambda)$은 어떻게 계산할 수 있는가?

문제 2(디코딩) : 주어진 연쇄 O와 은닉 마르코프 모델 $\lambda = (A, B)$에

[1] 여기서 A는 앞의 은닉 마르코프 정의에 의해 상태들 간의 전이 확률을 나타내며, B는 관찰되는 대상의 특정 상태에서의 방출 확률이다.



$\alpha_1(C) = .7$
$\alpha_2(C) = .7 \times .15 + .3 \times .08 = .129$
$\alpha_3(C) = .0129 \times .15 + .281 \times .32 = .10927$
$\alpha_1(H) = .3$
$\alpha_2(H) = .3 \times .12 + .7 \times .35 = .281$
$\alpha_3(C) = .281 \times .48 + .0129 \times .35 = .18003$

(1) $P(e|s)P(C|s) = 1 \times .7$
(2) $P(e|s)P(H|s) = 1 \times .3$
(3) $P(1|C)P(C|C) = .5 \times .3$
(4) $P(1|C)P(H|C) = .5 \times .7$
(5) $P(1|H)P(H|H) = .2 \times .6$
(6) $P(1|H)P(C|H) = .2 \times .4$

그림 6.6 아이스크림 수 1212 연쇄의 순방향 격자

로 전이할 확률 $P(1|H)P(C|H)$ 을 곱한다. 이제 이 두 확률값을 더하여 $\alpha_2(C)$ 에 저장한다. 만일 α 를 설정하지 않는다면 매 단계별로 앞에서 계산했던 과정을 반복해야 한다. 이런 식으로 각 단계별 값을 저장하는 $\alpha_t(j)$ 는 다음과 같이 형식화될 수 있다.

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_i(o_t) \tag{6.5}$$

즉, 순방향 매개변수 $\alpha_t(j)$ 는 그 전에 계산된 α_{t-1} 값에 i 에서 j 상태로 들어오는 전이 확률 (a_{ij}) 과 i 상태에서 방출되는 확률 $b_i(o_t)$ 값들을

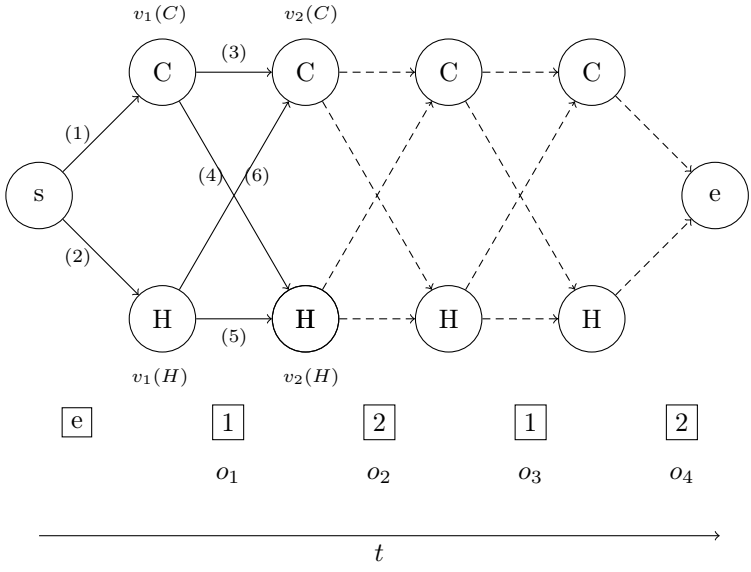
하여 더하거나, 역방향으로 뒤에서부터 첫 관찰 대상에 이르기까지의 β 값을 계산하여 구할 수 있다. 역방향 계산에서는 관찰 대상의 끝에서부터 첫 글자, β_1 에 이른 후 여기에 다시 시작 확률 (π) 을 곱하여 전체 연쇄의 확률값을 구해야 한다. 이제 아이스크림 수 1212 의 역방향으로 계산된 결과를 표 6.6 에서 살펴보자.

표 6.6 역방향에 의한 abab 의 확률값

관찰 대상	e	2	12	212	1212
$\beta(C)$	1	0.5	0.355	0.10085	0.0777355
$\beta(H)$	1	0.8	0.136	0.17888	0.0295336

역방향은 관찰되는 연쇄의 역순으로 계산이 이루어진다. 따라서 2 가 먼저 계산된다. 여기서는 관찰 대상의 끝에서는 반드시 종결 (end) 상태로 전이해야 하고 이때 빈 글자가 방출되는 것으로 가정한다. 그래서 빈 글자에 대한 확률 1 로 시작한다. 여기서 주의해야 할 것은 역방향 관점에서 이해되는 전이와 방출 확률이다. 예로 212 의 $\beta(C)$ 의 값이 계산되는 과정을 살펴보자. 우선 이전에 계산되어 온 12 의 $\beta(C)$ 0.355 에 $C \to C$ 전이 확률 0.3 과 C 에서 2 를 방출할 확률 0.5 를 곱한 것과 12 의 $\beta(H)$ 의 0.136 에 $C \to H$ 의 전이 확률 0.7 과 C 에서 2 를 방출할 확률 0.5 를 곱한 것을 서로 더하게 된다. 여기서 12 의 $\beta(H)$ 에서 순방향이라면 $H \to C$ 의 전이 확률을 곱해야 하지만 (6.7) 에서 살펴본 대로, 역방향이기 때문에 순방향으로는 C 에서 H 로 전이되는 것이므로 $C \to H$ 의 확률을 곱하게 된다.

최종적으로 계산된 $\beta(H)$ 와 $\beta(C)$ 를 더하면 표 6.2 에서 순방향으로 계산된 확률값과 같지 않다. 왜냐하면 수식 (6.8) 에서 보듯이 역방향에서는 시작 확률을 최종 β_1 에 곱해 주어야 하기 때문이다. 이제 시작 상태에서 C 와 H 에 이르는 확률을 β_1 에 곱하고 이 두 값을 더하면



$v_1(C) = .7$
$v_2(C) = \max(.7 \times .15, .3 \times .08) = .105$

$v_1(H) = .3$
$v_2(H) = \max(.3 \times .12, .7 \times .35) = .245$

(1) $P(e|s)P(C|s) = 1 \times .7$
(2) $P(e|s)P(H|s) = 1 \times .3$
(3) $P(1|C)P(C|C) = .5 \times .3$
(4) $P(1|C)P(H|C) = .5 \times .7$
(5) $P(1|H)P(H|H) = .2 \times .6$
(6) $P(1|H)P(C|H) = .2 \times .4$

그림 6.8 1212의 최적의 확률을 찾기 위한 Viterbi 격자

을 곱하여 구한 0.105와, $H \rightarrow C$ 전이의 경우인 $P(C|H)$과 $P(1|H)$을 곱한 후 다시 이전 $v_1(H)$ 값 0.3을 곱한 0.024를 비교한다. 둘 중에서 더 큰 값인 $C \rightarrow C$에서 전이되는 0.105를 취하게 된다. 이런 방법으로 각 상태에서 최대의 값을 v에 저장한 후 최종적으로 가장 큰 확률값을 가지는 노드들의 연쇄를 구하게 된다.

이제 1212의 최적의 연쇄를 구하는 전 과정을 살펴보자. 표 6.8은 Viterbi 알고리즘에 의한 이 연쇄의 확률값을 단계별로 정리한 것이다.

표 6.8은 상태 C, H의 Viterbi 값을 저장하는 v와 각 상태로 들어오는 확률값들을 보여 주고 있다. 첫 $v_1(C)$는 0.7이고 $v_1(H)$는 0.3이다.

표 6.8 Viterbi 알고리즘에 의한 1212 의 확률값

상태 연쇄	e	1	12	121	1212
$v(C)$	0.7	0.105	0.0784	0.01176	0.0087808
$v(H)$	0.3	0.245	0.1176	0.02744	0.0131712
	C	$C \to C$	$C \to C$	$C \to C$	$C \to C$
	0.7	0.105	0.01575	0.01176	0.001764
	H	$H \to C$	$H \to C$	$H \to C$	$H \to C$
	0.3	0.024	0.0784	0.009408	0.0087808
		$H \to H$	$H \to H$	$H \to H$	$H \to H$
		0.036	0.1176	0.014112	0.0131712
		$C \to H$	$C \to H$	$C \to H$	$C \to H$
		0.245	0.03675	0.02744	0.004116

출력 연쇄에 따라 C 상태로 들어오는 전이 $C \to C, H \to C$ 와 H 상태로 들어오는 $C \to H, H \to H$ 를 앞에서 설명한 대로 계산한 후 최대값을 $v(C), v(H)$ 에 저장하여 최종 연쇄에 이르게 된다. 이 경우 1212 의 최적 확률값은 0.0131712 이다. 전체적으로 최적의 연쇄만을 따라오면 표 6.9 와 같이 전체 연쇄를 알 수 있다.

이제 아이스크림 수 연쇄 1212 의 숨겨진 상태 연쇄 중에서 가장 확률이 높은 CHCHH 로 우리는 2009 년 여름의 날씨를 추정할 수 있게 된다. 여기서는 시작에서 빈 숫자를 가정했기 때문에 실제적으로는 e1212 연쇄에 대한 최대의 확률 연쇄가 된다.

Viterbi 알고리즘은 이렇게 각 상태마다 최대 확률값을 갖는 전이만을 저장한 후 다음 단계는 이 최적의 상태를 따르게 한다. 따라서 최대의 확률값을 갖지 않는 연쇄는 계산을 하지 않는다. 이 예에서 총 $2^5 = 32$ 가지의 가능한 연쇄 중에서 처음부터 최댓값을 갖는 연쇄의 확률값만을

표 6.9 아이스크림 수 연쇄 1212의 전체 상태 연쇄

상태 연속	e	1	12	121	1212
$v(C)$	0.7	0.105	0.0784	0.01176	0.0087808
$v(H)$	0.3	0.245	0.1176	0.02744	0.0131712
	C	CC	CCC	$CHCC$	$CHCCC$
	0.7	0.105	0.01575	0.01176	0.001764
	H	HC	CHC	$CHHC$	$CHCHC$
	0.3	0.024	0.0784	0.009408	0.0087808
		HH	CHH	$CHHH$	$CHCHH$
		0.036	0.1176	0.014112	0.0131712
		CH	CCH	$CHCH$	$CHCCH$
		0.245	0.03675	0.02744	0.004116

계산하여 진행하기 때문에 불필요한 연산을 피할 수 있다.

6.3.3 문제 3: 은닉 마르코프 모델의 학습

은닉 마르코프 모델과 관련한 마지막 문제는 매개변수인 전이 확률과 방출 확률을 어떻게 학습(training)할 수 있는가에 대한 것이다. 앞에서의 예는 관찰 연쇄 1212에 대한 각 상태에서의 전이 확률과 방출 확률이 주어진 상태에서 그 연쇄에 대한 확률과 최적의 확률값을 구하는 과정에 대한 것이었다. 그럼 이 전이 확률과 방출 확률은 어떻게 구할 수 있는지 생각해 보자.

우선 마르코프 연쇄에서 이런 매개변수를 학습하는 과정에 대해 살펴보자. 마르코프 연쇄는 상태 연쇄들이 숨겨져 있지 않고 그대로 드러나 있기 때문에 특정 관찰 연쇄에 대해 어떤 상태 연쇄를 따라야 하는지를 직접 알 수 있다. 앞 6.1.1의 Mealy 기계와 같은 경우 한 상태에서 다른

3:379-440.

Jeffreys, H. (1948), *Theory of Probability*, Clarendon Press, Oxford.

Jelinek, F., and R. L. Mercer (1980), "Interpolated estimation of Markov source parameters from sparse data," In *Proceedings of the Workshop on Pattern Recognition in Practice*, Amsterdam, The Netherlands.

Jurafsky, D., and J. H. Martin (2008), *Speech and Language Processing: An Introduction to Natural Language Processing and Computational Linguistics, and Speech Recognition*, 2nd Edition, Pearson Education International.

Karttunen, L. (1983), "Kimmo: a general morphological processor," *Texas Linguistics Forum*, 16:423-431.

Katz, S. M. (1987), "Estimation of probabilities from sparse data for the language model component of speech recognizer," *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35-3:400-401.

Kneser, R., and H. Ney (2005), "Improved backing-off for N-gram language modeling," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1:181-184.

Ko, H., A. Perovic, T. Ionin, and K. Wechsler (2006), "Adult L2-learners lack the Maximality Presuppostion, too". In K. U. Deen et. al (eds), *The Proceedings of the Inaugural Conference on Generative Approches to Language Acquitision*, North America, Honolulu, HI.

Lidstone, G.J. (1920), "Note on the general case of the Bayes-Laplace Formula for inductive or a posteriori probabilities," *Transactions of the Faculty of Actuaries*, 8:182-192.

Manning, C., and H. Schütze (1999), *Foundations of Statistical Natural Language Processing*, MIT Press.

Minium, E. W., B. M. King, and G. Bear (1993), *Statistical Reasoning in Psychology and Education*, Third Edition, John Wiley & Sons.

Minium, E. W., R. C. Clarke, and T. Coladarci (1998), *Elements of Statistical Reasoning*, John Wiley & Sons.

Ney, H., U. Essen, and R. Kneser (1994), "On structuring probabilistic dependencies in stochastic language modeling," *Computer, Speech, and Language*, 8:1-38.

Nugues, P. M. (2006), *An Introduction to Language Processing with Perl and Prolog*, Springer.

Quinlan, J. R. (1986), "Induction of decision trees," *Machine Learning*, 1-1:81-106.

Rabiner, L. R. (1989), "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77-2:257-286.