

Lecture 2: Delta Method

March, 2026

Lecturer: Jong-June Jeon

Scribe: Jong-June Jeon

1 Delta Method는 왜 필요한가?

모수 θ 의 추정량 T_n 이 다음을 만족한다고 가정하자:

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} Z$$

우리는 $\phi(\theta)$ 에 관심이 있으며, 그 자연스러운 추정량은 $\phi(T_n)$ 이다. $\phi(T_n)$ 에 대한 점근적 근사를 할 수 있을까? 이런 고민은 어디에서 출발했는지 살펴보자. 다음은 표본 상관 계수의 점근적 추론에 관하여이다.

Example 1.1. $i = 1, \dots, n$ 에 대해 (X_i, Y_i) 가 *i.i.d.*라고 하자. $\theta = (\mathbf{E}X, \mathbf{E}Y, \mathbf{E}X^2, \mathbf{E}Y^2, \mathbf{E}XY)^\top$ 라고 하면 자연스러운 θ 에 대한 추정값 T_n 은

$$T_n = (\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY})$$

을 생각할 수 있다. T_n 의 원소들은 독립 확률변수의 합의 형태이므로 $\mathbf{E}X^4 < \infty$ 와 $\mathbf{E}Y^4 < \infty$ 면

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} \text{Normal}$$

이 됨을 예상할 수 있다. T_n 의 공분산을 계산함으로써 중심극한정리의 공분산 항을 계산할 수 있다. 하지만 만약 우리가 관심있는 항이

$$\rho = \frac{\mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y}{\sqrt{\mathbf{E}X^2 - (\mathbf{E}X)^2}\sqrt{\mathbf{E}Y^2 - (\mathbf{E}Y)^2}}$$

라면 표본상관계수 $\hat{\rho}$ 에 대한 점근적 분석이 가능할까?

$$\hat{\rho} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sqrt{\overline{X^2} - (\bar{X})^2}\sqrt{\overline{Y^2} - (\bar{Y})^2}}$$

1) 확률수렴은 continuous mapping theorem 에 의해 쉽게 보일 수 있다. 하지만 $\hat{\rho}$ 의 정규근사는 자명한 것은 아니다. Delta Method 는 미분가능한 함수의 변환으로 정의된 새로운 estimator의 점근적 분석에 유용하게 활용된다. 그 아이디어는 테일러 전개에서 출발한다. $\theta \in \mathbb{R}$ 인 경우를 생각해본다. 먼저 $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ 을 증명하였다고 가정한다. 개략적으로

$$\phi(T_n) = \phi(\theta) + \phi'(\theta)(T_n - \theta) + C|T_n - \theta|^2$$

as $T_n - \theta \xrightarrow{p} 0$ 으로 볼 수 있다. 이는

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) = \phi'(\theta)\sqrt{n}(T_n - \theta) + C\sqrt{n}|T_n - \theta|^2$$

이 성립하고 $\sqrt{n}(T_n - \theta) = O_p(1)$ 고 $(T_n - \theta) = o_p(1)$ 이므로 $\sqrt{n}|T_n - \theta|^2 = o_p(1)$ 을 안다. Slutsky 정리에 의해 $\phi'(\theta)\sqrt{n}(T_n - \theta) \xrightarrow{d} \phi'(\theta)N(0, \sigma^2)$ 는 $\sqrt{n}(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \phi'(\theta)N(0, \sigma^2)$ 를 의미한다.

2 Multivariate Delta Method

이제 Delta Method를 다차원 벡터의 근사에 적용해 보겠다. 앞서 살펴본 것처럼 Delta Method의 핵심적인 역할을 하는 것은 미분이다. 다차원에서 사용하는 미분을 정의한다. $T_n \in \mathbb{R}^k$ 가 모수 $\theta \in \mathbb{R}^k$ 의 추정량이라 하자. 함수 $\phi: \mathbb{R}^k \rightarrow \mathbb{R}^m$ 에 대해 관심 있는 값은 $\phi(\theta)$ 이며, 자연스러운 추정량은 $\phi(T_n)$ 이다. 우리는 다음과 같은 점근정규성을 가정한다:

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} T.$$

ϕ 가 θ 에서 미분가능하다는 것은 어떤 선형사상 $\phi'_\theta: \mathbb{R}^k \rightarrow \mathbb{R}^m$ 가 존재하여

$$\phi(\theta + h) - \phi(\theta) = \phi'_\theta(h) + o(\|h\|) \quad (h \rightarrow 0)$$

를 만족하는 것을 의미한다. 여기서 $\|h\|$ 는 유클리드 노름이며, 모든 항은 \mathbb{R}^m 의 벡터이다. 일반적으로 ϕ 가 두 번 연속 미분가능(C^2)하면, 편미분들이 연속이므로 Jacobian D_θ 가 존재하여 $\phi'_\theta(h) = D_\theta h$ 가 된다.

Definition 2.1 (미분가능성). 함수 $\phi: \mathbb{R}^k \rightarrow \mathbb{R}^m$ 가 점 $\theta \in \mathbb{R}^k$ 에서 **미분가능(Differentiable)**하다는 것은 다음 조건을 만족하는 선형 사상(Linear map) $\phi'_\theta: \mathbb{R}^k \rightarrow \mathbb{R}^m$ 가 존재함을 의미한다:

$$\phi(\theta + h) - \phi(\theta) = \phi'_\theta(h) + o(\|h\|) \quad (h \rightarrow 0) \tag{1}$$

여기서 $\|h\|$ 는 유클리드 노름(Euclidean norm)이며, 식의 모든 항은 \mathbb{R}^m 의 벡터이다.

Example 2.2. 다음 함수를 생각하자:

$$\phi(x, y) = \begin{pmatrix} x^2 + y \\ xy \end{pmatrix}.$$

ϕ 가 두 번 연속 미분가능(C^2)이므로

$$\phi'_\theta(h) = \begin{pmatrix} 2x & 1 \\ y & x \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

Theorem 2.3. $\phi : D_\phi \subset \mathbb{R}^k \rightarrow \mathbb{R}^m$ 이 \mathbb{R}^k 의 부분집합 위에 정의된 함수이고, θ 에서 미분가능하다고 하자. 또한 T_n 이 ϕ 의 정의역 안에서 값을 가지는 확률벡터라고 하자. 수열 $r_n \rightarrow \infty$ 에 대하여 $r_n(T_n - \theta) \xrightarrow{d} T$ 라고 가정하자. 그러면

$$r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \phi'_\theta(T) = D_\theta T.$$

또한 다음이 성립한다:

$$r_n(\phi(T_n) - \phi(\theta)) - \phi'_\theta(r_n(T_n - \theta)) \xrightarrow{p} 0.$$

Proof. $r_n(T_n - \theta)$ 가 분포수렴하므로 Prohorov 정리에 의해 $r_n(T_n - \theta) = O_p(1)$ 이고 $T_n - \theta \xrightarrow{p} 0$ 이다. ϕ 가 θ 에서 미분가능하므로 나머지항 $R(h) = \phi(\theta + h) - \phi(\theta) - \phi'_\theta(h)$ 는 $h \rightarrow 0$ 일 때 $R(h) = o(\|h\|)$ 를 만족한다.

$$\phi(T_n) - \phi(\theta) - \phi'_\theta(T_n - \theta) = R(T_n - \theta) = o_p(\|T_n - \theta\|)$$

이 식의 양변에 r_n 을 곱하면

$$r_n(\phi(T_n) - \phi(\theta) - \phi'_\theta(T_n - \theta)) = o_p(r_n\|T_n - \theta\|).$$

$o_p(r_n\|T_n - \theta\|)$ 의 정의에 의해 $Y_n = o_p(1)$ 이 존재하여 $o_p(r_n\|T_n - \theta\|) = Y_n r_n\|T_n - \theta\|$ 를 만족하며 $r_n(T_n - \theta) = O_p(1)$ 이므로 $r_n\|T_n - \theta\| = O_p(1)$ 이다. 이 두 결과는 따라서 $o_p(r_n\|T_n - \theta\|) = o_p(1)$ 임을 의미한다. 이로부터 정리의 두 번째 주장, 즉 차이가 확률적으로 0으로 수렴함을 얻는다. 또한 선형사상 $D_\theta := \phi'_\theta$ 는 연속이므로 continuous mapping theorem에 의해 $D_\theta(r_n(T_n - \theta)) \xrightarrow{d} D_\theta(T)$ 이다. 마지막으로 Slutsky의 정리를 적용하면 $r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{d} D_\theta(T)$ 를 얻는다. \square

Example 2.4 (표본상관계수의 점근적 근사). (X_i, Y_i) 가 *i.i.d.* 라 하고 $g(x, y) = (x, y, x^2, y^2, xy)^\top \in \mathbb{R}^5$. 를 정의한다. 표본평균 벡터를

$$T_n = \frac{1}{n} \sum_{i=1}^n g(X_i, Y_i) = (\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY})^\top$$

로 둔다. 모멘트 벡터는 $\theta = \mathbb{E}[g(X, Y)] = (\mu_X, \mu_Y, m_{X^2}, m_{Y^2}, m_{XY})^\top$ 이며, 다변량 중심극한정리에 의해

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \Sigma), \quad \Sigma = \text{Cov}(g(X, Y)).$$

$\theta = (a, b, c, d, e) = (\mu_X, \mu_Y, m_{X^2}, m_{Y^2}, m_{XY})$ 라 하자.

$$A = c - a^2 = \text{Var}(X), \quad B = d - b^2 = \text{Var}(Y), \quad C = e - ab = \text{Cov}(X, Y).$$

모상관계수는 다음 함수로 표현된다:

$$\rho = \phi(\theta) = \frac{C}{\sqrt{AB}} = \frac{e - ab}{\sqrt{(c - a^2)(d - b^2)}}.$$

표본상관계수는

$$\hat{\rho} = \phi(T_n) = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sqrt{(\overline{X^2} - \bar{X}^2)(\overline{Y^2} - \bar{Y}^2)}}$$

로 주어진다. $\nabla\phi(\theta) = (\frac{\partial\phi}{\partial a}, \frac{\partial\phi}{\partial b}, \frac{\partial\phi}{\partial c}, \frac{\partial\phi}{\partial d}, \frac{\partial\phi}{\partial e})^\top$ 는 다음과 같은 각 편미분 값으로 계산된다:

$$\frac{\partial\phi}{\partial a} = -\frac{b}{\sqrt{AB}} + \rho\frac{a}{A}, \quad \frac{\partial\phi}{\partial b} = -\frac{a}{\sqrt{AB}} + \rho\frac{b}{B},$$

$$\frac{\partial\phi}{\partial c} = -\frac{\rho}{2A}, \quad \frac{\partial\phi}{\partial d} = -\frac{\rho}{2B}, \quad \frac{\partial\phi}{\partial e} = \frac{1}{\sqrt{AB}}.$$

따라서

$$\nabla\phi(\theta) = \left(-\frac{b}{\sqrt{AB}} + \rho\frac{a}{A}, -\frac{a}{\sqrt{AB}} + \rho\frac{b}{B}, -\frac{\rho}{2A}, -\frac{\rho}{2B}, \frac{1}{\sqrt{AB}} \right)^\top.$$

귀무가설 $H_0 : \rho = 0$ 하에서 점근적 분포를 계산해보자. $\rho = 0$ 일때 모수벡터를 θ_0 이라 표기한다.

$$\nabla\phi(\theta_0) = \left(-\frac{\mu_Y}{\sqrt{AB}}, -\frac{\mu_X}{\sqrt{AB}}, 0, 0, \frac{1}{\sqrt{AB}} \right)^\top.$$

따라서 델타 방법에 의해

$$\sqrt{n} \hat{\rho} \xrightarrow{d} N(0, V_0), \quad V_0 = \nabla \phi(\theta_0)^\top \Sigma \nabla \phi(\theta_0).$$

또한 V_0 는 $\text{Var}(AX) = A\text{Var}(X)A^\top$ 임을 이용하면

$$V_0 = \text{Var}\left(\nabla \phi(\theta_0)^\top g(X, Y)\right) = \frac{1}{AB} \text{Var}(XY - \mu_Y X - \mu_X Y).$$

그런데

$$(X - \mu_X)(Y - \mu_Y) = XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y$$

이므로 분산은 상수 이동에 불변이고, H_0 에서 $\mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = 0$ 이기 때문에

$$\text{Var}(XY - \mu_Y X - \mu_X Y) = \mathbb{E}[(X - \mu_X)^2(Y - \mu_Y)^2].$$

결과적으로

$$V_0 = \frac{\mathbb{E}[(X - \mu_X)^2(Y - \mu_Y)^2]}{\text{Var}(X)\text{Var}(Y)}$$

*Plug-in estimator*는 다음과 같이 주어진다.

$$\hat{A} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \hat{B} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2, \quad \hat{M} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2$$

로 두면 $\hat{V}_0 = \frac{\hat{M}}{\hat{A}\hat{B}}$ 를 얻는다.

3 분산 안정화 변환에 응용

3.1 Plug-in 추정량의 신뢰구간과 유한표본하에서 문제점

확률변수 수열 X_n 이 다음과 같은 점근적 정규성을 만족한다고 가정하자.

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$$

여기서 점근적 분산은 미지의 모수 θ 에 의존하고 있다. 예를 들어 X_1, \dots, X_n 이 모수 p 를 가진 베르누이 분포로부터의 iid 열이라고 하자. CLT에 의해 표본 비율 \hat{p}_n 은 다음을 만족한다:

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{d} N(0, p(1-p)) \tag{2}$$

이 예에서 점근적 분산 $\sigma^2(p) = p(1-p)$ 는 미지의 파라미터 p 에 의존함을 알 수 있다. Plug-in 방법을 적용하여 p 대신 일치 추정량 \hat{p}_n 을 대입하면, 추정된 점근적 분산은 다음과 같이 주어진다.

$$\hat{V} = \hat{p}_n(1 - \hat{p}_n)$$

함수 $\sigma(p) = \sqrt{p(1-p)}$ 가 $p \in (0, 1)$ 에서 연속이므로, 대수의 법칙(LLN)에 의한 $\hat{p}_n \xrightarrow{p} p$ 성질과 Slutsky's Theorem에 의해 다음의 수렴성이 성립한다.

$$\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{\hat{p}_n(1 - \hat{p}_n)}} \xrightarrow{d} N(0, 1)$$

즉, 모수 p 를 추정량 \hat{p}_n 으로 대체하더라도 분모가 확률 수렴하기 때문에 전체 통계량의 점근적 정규성은 여전히 유지된다. 이를 통해 계산된 $1 - \alpha$ 점근적 신뢰구간(일명 Wald 신뢰구간)은 아래와 같다.

$$CI_{1-\alpha} \approx \left[\hat{p}_n - z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}}, \quad \hat{p}_n + z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \right]$$

하지만 일반적인 Plug-in 방식은 점근적 정규성에 기반하여 다음과 같은 대칭적 신뢰구간을 생성한다. 이 방식은 \hat{p} 를 중심으로 좌우 거리가 동일하므로, p 가 0 또는 1에 가까운 비대칭 분포 상황에서도 강제로 대칭적인 구간을 설정하는 오류를 범할 수 있다. Figure 1은 유한표본 하에서 Plug-in 방식(예: Wald interval)은 신뢰구간이 파라미터의 정의구역(Space)을 벗어남을 보여준다. 이항분포에서 $p \approx 0$ 일 때, Wald interval은 하한값으로 음수를 가질 수 있다. Figure 2는 추정량의 분포를 보다 비대칭일 수 있는데 실제로 이를 반영하지 못하는 문제를 보여준다.

3.2 분산안정화 변환 (Variance Stabilizing Transformation)

앞서 살펴보았듯이 만약 $\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$ 이면 미분 가능한 함수 g 에 대하여 델타 방법(Delta Method)은 다음을 보장한다.

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, [g'(\theta)]^2 \sigma^2(\theta))$$

이를 이용하여 (2)에서 우리는 점근적 분산을 파라미터 p 와 무관하게 만들기 위해, $g'(p)\sigma(p) = C$ 를 만족하는 변환 함수 g 를 찾을 수 있다. 계산의 편의를 위해 $C = 1$ 로 설정하면:

$$g'(p)\sqrt{p(1-p)} = 1 \quad \rightarrow \quad g'(p) = \frac{1}{\sqrt{p(1-p)}}$$

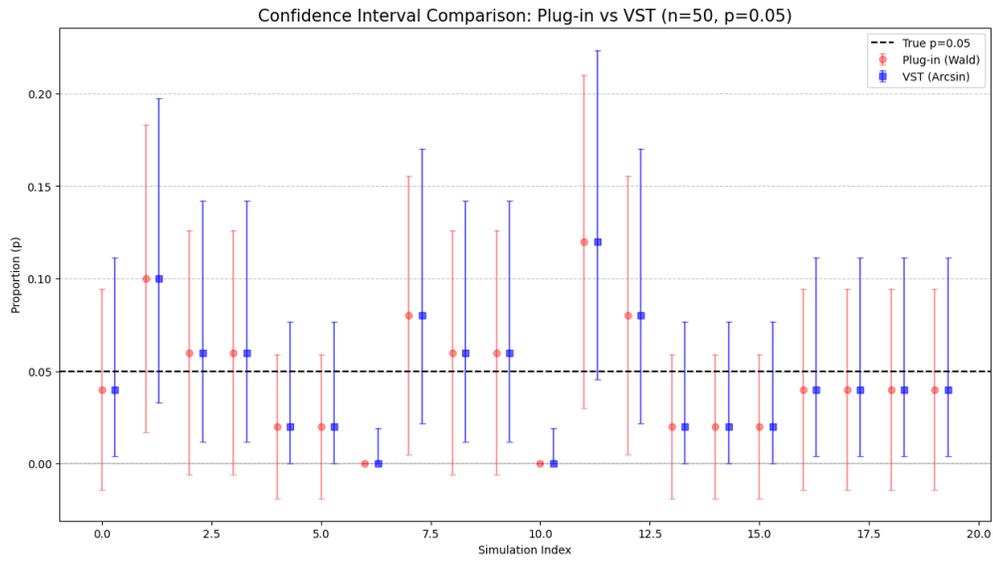


Figure 1: VST: 분산안정화 방법 적용 후 역변환하여 얻은 신뢰구간, Plug-in: 정규근하를 통해 얻는 신뢰구간

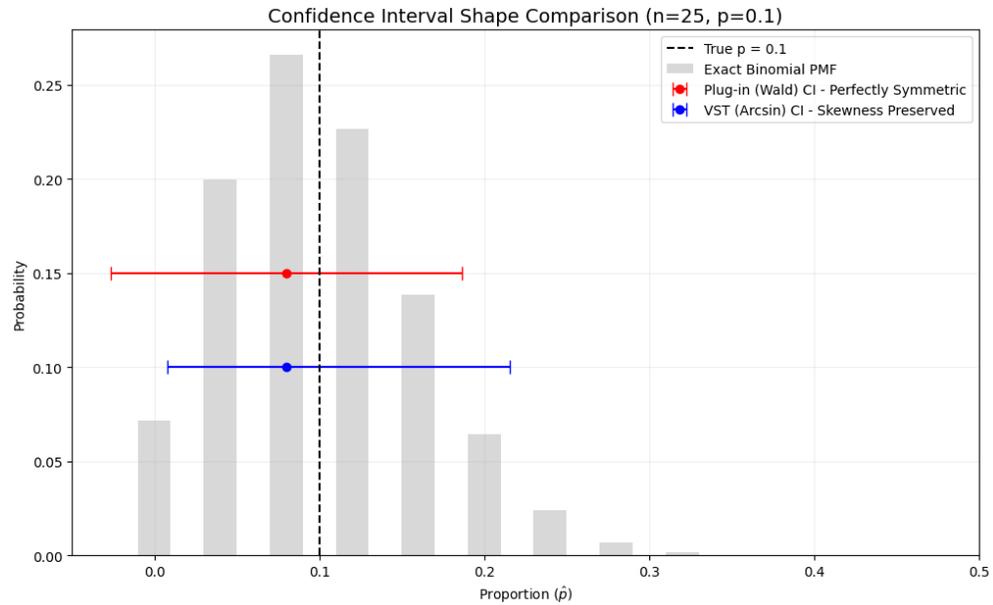


Figure 2: Exact Binomial PMF: 유한샘플에서 실제 \hat{p} 의 분포, VST: 분산안정화 방법 적용 후 역변환하여 얻은 신뢰구간, Plug-in: 정규근사를 통해 얻는 신뢰구간

이 미분방정식을 적분하면 다음과 같은 아크사인 변환(Arcsin Transformation)을 얻는다:

$$g(p) = 2 \arcsin(\sqrt{p})$$

이 변환을 델타 방법(Delta Method)을 적용하면, 변환된 통계량은 다음과 같이 분산이 1로 안정화된 점근적 정규분포를 따른다:

$$\sqrt{n}(g(\hat{p}_n) - g(p)) \xrightarrow{d} N(0, 1)$$

그러면 \hat{p} 에 대한 신뢰구간을 다음과 같이 얻을 수 있다.

- (1단계: 변환된 공간에서의 신뢰구간 구하기) 먼저 $g(p)$ 에 대한 점근적 $100(1 - \alpha)\%$ 신뢰구간을 구하면 다음과 같이 대칭적인 형태이다:

$$\left[2 \arcsin(\sqrt{\hat{p}_n}) - \frac{z_{\alpha/2}}{\sqrt{n}}, \quad 2 \arcsin(\sqrt{\hat{p}_n}) + \frac{z_{\alpha/2}}{\sqrt{n}} \right]$$

- (2단계: 원래 공간으로의 역변환) 위 구간의 양 끝점에 g 의 역함수인 $g^{-1}(y) = \sin^2(y/2)$ 를 적용하여 p 에 대한 최종 신뢰구간을 도출한다:

$$CI_{VST} = \left[\sin^2 \left(\arcsin(\sqrt{\hat{p}_n}) - \frac{z_{\alpha/2}}{2\sqrt{n}} \right), \quad \sin^2 \left(\arcsin(\sqrt{\hat{p}_n}) + \frac{z_{\alpha/2}}{2\sqrt{n}} \right) \right]$$

Figure 1와 Figure 2에서 확인할 수 있는 것 처럼 분산안정화 변환을 통해서 유한표본하에서 더 개선된 신뢰구간을 얻을 수 있다.

4 연습문제

- Example 1.1 에서 T_n 의 정규근사를 위해서 $\mathbf{E}X^4, \mathbf{E}Y^4 < \infty$ 가 필요한 이유는 무엇인가?
- 표분분산의 점근적 분포를 Delta Method 로 유도하고 이론적인 분포를 시뮬레이션을 통한 결과로 검증하여라.
- 분산안정화 변환이 필요한 예를 제시하고 시뮬레이션 결과로 주장을 뒷받침하여라.