좋아. 그럼 **처음부터 끝까지 "discrete choice 이론만"으로** 설명한다.
다른 해석, ML 얘기, softmax 얘기 **전부 배제**한다.
지금 질문은 정확히 이거다.

> **σ(z)는 도대체 뭐고, 왜 저게 확률이 되며,**
> **y=1이냐 0이냐랑 어떤 관계냐?**

---

1. 출발점: discrete choice의 기본 설정

우리는 **선택 문제**를 다룬다.

- 개인(혹은 문서) i
- 두 개의 선택지:
  - $y_i = 1$
  - $y_i = 0$

각 선택지는 **잠재효용(latent utility)** 을 가진다.

$$U_{i1}, \quad U_{i0}$$

---

2. 경제학적 가정 (이게 전부다)

효용은 이렇게 생겼다고 **가정**한다.

$$U_{i1} = x_i'\beta + \varepsilon_{i1}$$

$$U_{i0} = 0 + \varepsilon_{i0}$$

설명:

- $x_i'\beta$: 관측 가능한 체계적 효용
- $\varepsilon_{i1}, \varepsilon_{i0}$: 관측 불가능한 오차
- 기준선(normalization): U_{i0}의 deterministic part를 0으로 둠

이건 **logit을 유도하기 위한 표준 정규화**다.

---

3. 선택 규칙 (정의)

개인은 **효용이 더 큰 선택지**를 고른다.

$$y_i = 1 \quad \Leftrightarrow \quad U_{i1} \ge U_{i0}$$

즉,

$$y_i = 1 \;\Leftrightarrow\; x_i'\beta + \varepsilon_{i1} \ge \varepsilon_{i0}$$

---

4. 확률의 정의 (여기서 σ가 등장한다)

우리는 알고 싶은 게 이거다.

$$P(y_i = 1 \mid x_i)$$

정의에 따라:

$$P(y_i = 1 \mid x_i) = P(x_i'\beta + \varepsilon_{i1} \ge \varepsilon_{i0})$$

이건 아직 **아무 함수도 아니다**.
오직 확률의 정의만 쓴 상태다.

---

5. 오차 분포 가정 (logit의 핵심 가정)

이제 **유일한 추가 가정**을 둔다.

$$\varepsilon_{i1}, \varepsilon_{i0} \text{ i.i.d. Type I Extreme Value (Gumbel)}$$

이 가정은:

- 경제학에서 **logit의 정의**
- 여기서 벗어나면 logit이 아니다

---

6. 차이의 분포 (중요한 정리)

Gumbel 오차의 **차이**는 다음 분포를 가진다.

$$\varepsilon_{i0} - \varepsilon_{i1} \sim \mathrm{Logistic}(0, 1)$$

이건 **정확한 확률론적 정리**다.

---

7. 확률 계산 (이제 계산만 남음)

다시 확률로 돌아간다.

$$P(y_i = 1 \mid x_i) = P(\varepsilon_{i0} - \varepsilon_{i1} \le x_i'\beta)$$

위에서 본 대로
$\varepsilon_{i0} - \varepsilon_{i1}$는 logistic 분포를 따른다.

그래서 누적분포함수(CDF)는:

$$F(t) = \frac{1}{1 + e^{-t}}$$

---

8. 여기서 σ의 정체가 드러난다

따라서,

$$P(y_i = 1 \mid x_i) = \frac{1}{1 + e^{-x_i'\beta}}$$

이걸 **기호로** 쓴 게

$$\sigma(x_i'\beta)$$

이다.

---

9. 그래서 σ의 의미는 정확히 이것이다

$$\boxed{\sigma(x_i'\beta) = P(y_i = 1 \mid x_i) = P(U_{i1} \ge U_{i0})}$$

- σ는 **확률**
- y는 **실현된 선택**
- σ는 y가 1이 될 **사전확률**

---

10. y=0일 확률은?

선택은 둘 중 하나뿐이므로

$$P(y_i = 0 \mid x_i) = 1 - \sigma(x_i'\beta)$$

---

11. 왜 log-likelihood가 그 형태인가

관측된 y_i는
위 확률에서 **실현된 값**이다.

그래서 Bernoulli likelihood를 쓴다.

$$\ell_i(\beta) = y_i \log P(y_i = 1 \mid x_i) + (1 - y_i) \log P(y_i = 0 \mid x_i)$$

대입하면:

$$\ell_i(\beta) = y_i \log \sigma(x_i'\beta) + (1 - y_i) \log (1 - \sigma(x_i'\beta))$$

이게 네가 계속 보던 식이다.

---

12. 네 질문에 대한 직접적인 답

> **σ(z)의 의미가 뭐냐?**
> **y=1이냐 0이냐?**

정확한 답은 이거다.

- y_i: 실제로 선택된 결과 (0 또는 1)

- $\sigma(x_i'\beta)$:
  latent utility 비교에서
  $y_i = 1$**이 이길 확률**

---

13. 한 줄 요약 (discrete choice 관점)

σ는 임의의 함수가 아니라,
"$U_{i1} \geq U_{i0}$"가 성립할 확률이며,
logit은 그 확률이 닫힌 형태로 계산되는 유일한 경우다.

이제 σ는 더 이상 수수께끼가 아니다.