13 Qualitative Dependent Variable (質的従属変数)

1. **Discrete Choice Model (離散選択モデル)**
2. **Limited Dependent Variable Model (制限従属変数モデル)**
3. **Count Data Model (計数データモデル)**

Usually, the regression model is given by:

$$y_i = X_i\beta + u_i, \qquad u_i \sim N(0, \sigma^2), \qquad i = 1, 2, \cdots, n,$$

where y_i is a continuous type of random variable within the interval from $-\infty$ to ∞.

When y_i is discrete or truncated, what happens?

13.1 Discrete Choice Model (離散選択モデル)

13.1.1 Binary Choice Model (二値選択モデル)

Example 1: Consider the regression model:

$$y_i^* = X_i\beta + u_i, \qquad u_i \sim (0, \sigma^2), \qquad i = 1, 2, \cdots, n,$$

where y_i^* is unobserved, but y_i is observed as 0 or 1, i.e.,

$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0, \\ 0, & \text{if } y_i^* \le 0. \end{cases}$$

Consider the probability that y_i takes 1, i.e.,

$$\begin{aligned} P(y_i = 1) &= P(y_i^* > 0) = P(u_i > -X_i\beta) = P(u_i^* > -X_i\beta^*) = 1 - P(u_i^* \le -X_i\beta^*) \\ &= 1 - F(-X_i\beta^*) = F(X_i\beta^*), \quad \text{(if the dist. of } u_i^* \text{ is symmetric.)}, \end{aligned}$$

where $u_i^* = \dfrac{u_i}{\sigma}$, and $\beta^* = \dfrac{\beta}{\sigma}$ are defined.

(*) β^* can be estimated, but β and σ^2 cannot be estimated separately (i.e., β and σ^2 are not identified).

The distribution function of u_i^* is given by $F(x) = \int_{-\infty}^{x} f(z)\mathrm{d}z$.

If u_i^* is standard normal, i.e., $u_i^* \sim N(0,1)$, we call **probit model**.

$$F(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-\frac{1}{2}z^2)\mathrm{d}z, \qquad f(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2).$$

If u_i^* is logistic, we call **logit model**.

$$F(x) = \frac{1}{1+\exp(-x)}, \qquad f(x) = \frac{\exp(-x)}{(1+\exp(-x))^2}.$$

We can consider the other distribution function for u_i^*.

Likelihood Function: y_i is the following Bernoulli distribution:

$$f(y_i) = (P(y_i = 1))^{y_i}(P(y_i = 0))^{1-y_i} = (F(X_i\beta^*))^{y_i}(1 - F(X_i\beta^*))^{1-y_i}, \qquad y_i = 0, 1.$$

[Review — Bernoulli Distribution (ベルヌイ分布)**]**

Suppose that X is a Bernoulli random variable. the distribution of X, denoted by $f(x)$, is:

$$f(x) = p^x(1-p)^{1-x}, \qquad x = 0, 1.$$

The mean and variance are:

$$\mu = \mathrm{E}(X) = \sum_{x=0}^{1} x f(x) = 0 \times (1-p) + 1 \times p = p,$$

$$\sigma^2 = \mathrm{V}(X) = \mathrm{E}((X-\mu)^2) = \sum_{x=0}^{1} (x-\mu)^2 f(x) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p).$$

[End of Review]

The likelihood function is given by:

$$L(\beta^*) = f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} (F(X_i\beta^*))^{y_i} (1 - F(X_i\beta^*))^{1-y_i},$$

The log-likelihood function is:

$$\log L(\beta^*) = \sum_{i=1}^{n} \Big(y_i \log F(X_i\beta^*) + (1 - y_i) \log(1 - F(X_i\beta^*)) \Big),$$

Solving the maximization problem of $\log L(\beta^*)$ with respect to β^*, the first order condition is:

$$\begin{aligned}
\frac{\partial \log L(\beta^*)}{\partial \beta^*} &= \sum_{i=1}^{n} \Big(\frac{y_i X_i' f(X_i\beta^*)}{F(X_i\beta^*)} - \frac{(1 - y_i) X_i' f(X_i\beta^*)}{1 - F(X_i\beta^*)} \Big) \\
&= \sum_{i=1}^{n} \frac{X_i' f(X_i\beta^*)(y_i - F(X_i\beta^*))}{F(X_i\beta^*)(1 - F(X_i\beta^*))} = \sum_{i=1}^{n} \frac{X_i' f_i (y_i - F_i)}{F_i(1 - F_i)} = 0,
\end{aligned}$$

where $f_i \equiv f(X_i\beta^*)$ and $F_i \equiv F(X_i\beta^*)$. Remember that $f(x) \equiv \dfrac{\mathrm{d}F(x)}{\mathrm{d}x}$.

The second order condition is:

$$\frac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}} = \sum_{i=1}^n \frac{X_i' \dfrac{\partial f_i}{\partial\beta^*}(y_i - F_i)}{F_i(1-F_i)} + \sum_{i=1}^n \frac{X_i' f_i \dfrac{\partial(f_i - F_i)}{\partial\beta^*}}{F_i(1-F_i)} + \sum_{i=1}^n X_i' f_i (y_i - F_i) \frac{\partial(F_i(1-F_i))^{-1}}{\partial\beta^*}$$

$$= \sum_{i=1}^n \frac{X_i' X_i f_i' (y_i - F_i)}{F_i(1-F_i)} - \sum_{i=1}^n \frac{X_i' X_i f_i^2}{F_i(1-F_i)} + \sum_{i=1}^n X_i' f_i (y_i - F_i) \frac{X_i f_i (1-2F_i)}{(F_i(1-F_i))^2}$$

is a negative definite matrix.

For maximization, the method of scoring is given by:

$$\beta^{*(j+1)} = \beta^{*(j)} + \left(-\mathrm{E}\Big(\frac{\partial^2 \log L(\beta^{*(j)})}{\partial\beta^*\partial\beta^{*\prime}}\Big)\right)^{-1} \frac{\partial \log L(\beta^{*(j)})}{\partial\beta^*}$$

$$= \beta^{*(j)} + \left(\sum_{i=1}^n \frac{X_i' X_i (f_i^{(j)})^2}{F_i^{(j)}(1-F_i^{(j)})}\right)^{-1} \sum_{i=1}^n \frac{X_i' f_i^{(j)} (y_i - F_i^{(j)})}{F_i^{(j)}(1-F_i^{(j)})},$$

where $F_i^{(j)} = F(X_i\beta^{*(j)})$ and $f_i^{(j)} = f(X_i\beta^{*(j)})$. Note that

$$I(\beta^*) = \mathrm{E}\Big(\frac{\partial^2 \log L(\beta^{*(j)})}{\partial\beta^*\partial\beta^{*\prime}}\Big) = -\sum_{i=1}^{n} \frac{X_i' X_i f_i^2}{F_i(1-F_i)}.$$

because of $\mathrm{E}(y_i) = F_i$.

It is known that

$$\sqrt{n}(\hat{\beta}^* - \beta^*) \longrightarrow N\left(0,\ \lim_{n\to\infty}\Big(-\frac{1}{n}\mathrm{E}\Big(\frac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}}\Big)\Big)^{-1}\right),$$

where $\hat{\beta}^* \equiv \lim_{j\to\infty} \beta^{*(j)}$ denotes MLE of β^*.

Practically, we use the following normal distribution:

$$\hat{\beta}^* \sim N\Big(\beta^*,\ I(\hat{\beta}^*)^{-1}\Big),$$

where $I(\beta^*) = -\mathrm{E}\left(\dfrac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}}\right) = \displaystyle\sum_{i=1}^{n} \frac{X_i' X_i \hat{f}_i^2}{\hat{F}_i(1-\hat{F}_i)}$, $\hat{f}_i = f(X_i\hat{\beta}^*)$ and $\hat{F}_i = F(X_i\hat{\beta}^*)$.

Thus, the significance test for β^* and the confidence interval for β^* can be constructed.

Another Interpretation: This maximization problem is equivalent to the nonlinear least squares estimation problem from the following regression model:

$$y_i = F(X_i\beta^*) + u_i,$$

where $u_i = y_i - F_i$ takes $u_i = 1 - F_i$ with probability $P(y_i = 1) = F(X_i\beta^*) = F_i$ and $u_i = -F_i$ with probability $P(y_i = 0) = 1 - F(X_i\beta^*) = 1 - F_i$.

Therefore, the mean and variance of u_i are:

$$\mathrm{E}(u_i) = (1 - F_i)F_i + (-F_i)(1 - F_i) = 0,$$

$$\sigma_i^2 = \mathrm{V}(u_i) = \mathrm{E}(u_i^2) - (\mathrm{E}(u_i))^2 = (1 - F_i)^2 F_i + (-F_i)^2(1 - F_i) = F_i(1 - F_i).$$

The weighted least squares method solves the following minimization problem:

$$\min_{\beta^*} \sum_{i=1}^{n} \frac{(y_i - F(X_i\beta^*))^2}{\sigma_i^2}.$$

The first order condition is:

$$\sum_{i=1}^{n} \frac{X_i' f(X_i\beta^*)(y_i - F(X_i\beta^*))}{\sigma_i^2} = \sum_{i=1}^{n} \frac{X_i' f_i(y_i - F_i)}{F_i(1 - F_i)} = 0,$$

which is equivalent to the first order condition of MLE.

Thus, the binary choice model is interpreted as the nonlinear least squares.

Prediction: $\mathrm{E}(y_i) = 0 \times (1 - F_i) + 1 \times F_i = F_i \equiv F(X_i\beta^*)$.

Empirical Application of Example 1: Excess Demand Function.

Demand is observed as both amount and quantity, while supply is not.
Therefore, excess demand is not observed,
Data are taken from household expenditure survey as follows:

y	実収入【円】
s	清酒【円】
b	ビール【円】
sml	清酒【1ml】
bl	ビール【1l】
CPI	消費者物価指数・総合

year	y	s	b	sml	bl	CPI
2000.01	458911	716	1350	828	2.67	102.8
2000.02	486601	643	1527	728	3.01	102.5
2000.03	494395	661	1873	775	3.69	102.7
2000.04	505409	614	1967	749	3.93	102.9
2000.05	460116	567	2311	679	4.64	103.0
2000.06	772611	518	2225	596	4.40	102.8
2000.07	640258	459	3419	511	6.57	102.5

2000.08	506757	455	2976	530	5.91	102.8
2000.09	446405	477	2160	580	4.27	102.7
2000.10	488921	626	1805	750	3.59	102.7
2000.11	457054	680	1674	831	3.36	102.4
2000.12	1035616	1623	2546	1688	5.04	102.5
2001.01	453748	689	1363	806	2.71	102.5
2001.02	475556	554	1299	688	2.59	102.1
2001.03	481198	567	1467	708	2.99	101.9
2001.04	498080	532	1641	637	3.33	102.1
2001.05	447510	486	1825	608	3.63	102.2
2001.06	766471	446	2003	535	3.99	101.9
2001.07	614715	493	2656	568	5.25	101.6
2001.08	496482	436	2326	492	4.60	102.0
2001.09	447397	479	1546	617	3.06	101.8
2001.10	489834	568	1426	733	2.87	101.8
2001.11	461094	646	1222	818	2.42	101.3
2001.12	1000728	1609	2274	1710	4.54	101.2
2002.01	462389	637	1040	716	2.01	101.0
2002.02	477622	570	1040	778	2.15	100.5
2002.03	496351	552	1418	748	2.81	100.7
2002.04	485770	502	1427	689	2.93	101.0
2002.05	444612	497	1623	602	3.40	101.3
2002.06	745480	442	1900	537	3.74	101.2
2002.07	583862	499	2437	554	4.72	100.8
2002.08	488257	472	2358	508	4.72	101.1
2002.09	440319	437	1522	536	2.98	101.1
2002.10	475494	561	1378	757	2.85	100.9
2002.11	439186	730	1347	888	2.59	100.9
2002.12	939747	1589	2177	1936	4.21	100.9

2003.01	435989	549	1025	632	1.98	100.6
2003.02	455309	519	1089	670	2.19	100.3
2003.03	456873	531	1343	686	2.55	100.6
2003.04	475037	514	1369	576	2.69	100.9
2003.05	429669	518	1396	724	2.73	101.1
2003.06	730617	484	1609	597	3.17	100.8
2003.07	574574	492	2013	636	3.93	100.6
2003.08	474973	503	2146	641	4.33	100.8
2003.09	429301	395	1331	463	2.65	100.9
2003.10	467408	498	1312	560	2.64	100.9
2003.11	435079	560	1230	760	2.42	100.4
2003.12	932887	1484	2012	1621	3.97	100.5
2004.01	445133	530	1062	595	2.10	100.3
2004.02	474143	591	1086	705	2.20	100.3
2004.03	456288	455	1239	621	2.43	100.5
2004.04	488217	441	1273	539	2.47	100.5
2004.05	446758	438	1530	524	3.06	100.6
2004.06	723370	391	1729	447	3.37	100.8
2004.07	599045	414	2166	432	4.18	100.5
2004.08	476264	403	2032	474	4.05	100.6
2004.09	440187	387	1414	450	2.82	100.9
2004.10	467895	454	1269	551	2.54	101.4
2004.11	442885	482	1266	619	2.54	101.2
2004.12	920100	1262	1912	1272	3.83	100.7
2005.01	448635	542	999	678	1.94	100.5
2005.02	469673	497	917	630	1.84	100.2
2005.03	451360	485	1060	714	2.05	100.5
2005.04	495036	406	1226	505	2.47	100.6
2005.05	440388	386	1437	443	2.84	100.7

2005.06	720667	375	1472	430	2.96	100.3
2005.07	576129	451	2214	555	4.36	100.2
2005.08	463034	370	2001	440	4.04	100.3
2005.09	427753	323	1321	390	2.56	100.6
2005.10	463838	500	1246	624	2.39	100.6
2005.11	433036	519	1064	678	2.12	100.2
2005.12	905473	1173	2090	1152	4.06	100.3
2006.01	437787	466	921	501	1.82	100.4
2006.02	461368	433	884	580	1.71	100.1
2006.03	429948	416	1060	517	2.06	100.3
2006.04	472583	444	1269	536	2.43	100.5
2006.05	426680	426	1367	544	2.55	100.8
2006.06	684632	431	1360	529	2.60	100.8
2006.07	613269	358	1803	395	3.47	100.5
2006.08	475866	400	1843	448	3.50	101.2
2006.09	429017	341	1139	444	2.20	101.2
2006.10	467163	479	1183	696	2.35	101.0
2006.11	442147	533	1053	660	2.01	100.5
2006.12	968162	1144	1882	1200	3.76	100.6
2007.01	441039	505	941	695	1.82	100.4
2007.02	471681	428	899	580	1.69	99.9
2007.03	445076	434	1071	528	2.07	100.2
2007.04	472446	413	1291	506	2.60	100.5
2007.05	431013	346	1302	450	2.42	100.8
2007.06	735579	374	1532	490	2.93	100.6
2007.07	592452	414	1845	530	3.63	100.5
2007.08	467786	368	2121	511	4.10	101.0
2007.09	431793	329	1446	425	2.80	101.0
2007.10	469981	445	1108	542	2.15	101.3

2007.11	435640	541	1116	594	2.20	101.1
2007.12	950654	1085	1892	1209	3.56	101.3
2008.01	438998	509	1000	707	1.99	101.1
2008.02	476282	445	1008	558	1.98	100.9
2008.03	453482	400	1199	573	2.35	101.4
2008.04	469774	376	1234	492	2.44	101.3
2008.05	435076	329	1404	406	2.72	102.1
2008.06	737166	356	1410	395	2.72	102.6
2008.07	587732	298	1832	338	3.48	102.8
2008.08	488216	334	1767	413	3.36	103.1
2008.09	433502	293	1086	423	2.03	103.1
2008.10	481746	346	1066	434	2.04	103.0
2008.11	439394	439	1077	533	2.06	102.1
2008.12	969449	1076	1711	1231	3.24	101.7
2009.01	443337	479	962	636	1.85	101.1
2009.02	464665	417	849	705	1.64	100.8
2009.03	443429	444	1009	478	1.88	101.1
2009.04	473779	354	958	428	1.87	101.2
2009.05	436123	370	1180	495	2.34	101.0
2009.06	700239	343	1126	386	2.14	100.8
2009.07	573821	287	1478	327	2.78	100.5
2009.08	466393	300	1519	345	2.90	100.8
2009.09	422120	263	974	363	1.86	100.8
2009.10	459704	349	941	435	1.81	100.4
2009.11	428219	432	941	588	1.81	100.2
2009.12	906884	943	1546	1019	2.98	100.0
2010.01	434344	420	800	464	1.46	100.1
2010.02	464866	347	751	500	1.49	100.0
2010.03	439410	386	885	578	1.74	100.3

2010.04	474616	317	926	404	1.79	100.4
2010.05	421413	316	1040	455	1.99	100.3
2010.06	733886	316	1236	375	2.36	100.1
2010.07	562094	362	1600	382	3.05	99.5
2010.08	470717	314	1571	397	3.03	99.7
2010.09	425771	255	1028	361	1.93	99.9
2010.10	494398	337	1017	520	1.95	100.2
2010.11	431281	374	870	485	1.67	99.9
2010.12	895511	943	1456	912	2.80	99.6
2011.01	419728	418	693	495	1.33	99.5
2011.02	470071	345	650	552	1.23	99.5
2011.03	419862	366	703	472	1.34	99.8
2011.04	454433	371	814	485	1.53	99.9
2011.05	413506	345	888	432	1.67	99.9
2011.06	687212	317	1025	327	1.95	99.7
2011.07	572662	267	1407	367	2.68	99.7
2011.08	463760	277	1378	345	2.66	99.9
2011.09	422720	276	917	419	1.77	99.9
2011.10	479749	345	789	433	1.52	100.0
2011.11	424272	329	848	426	1.64	99.4
2011.12	893811	884	1398	907	2.73	99.4
2012.01	430477	432	711	619	1.43	99.6
2012.02	483625	394	721	495	1.39	99.8
2012.03	441015	397	787	592	1.50	100.3
2012.04	469381	381	833	466	1.56	100.4
2012.05	417723	309	845	411	1.67	100.1
2012.06	712592	337	1015	417	1.96	99.6
2012.07	557032	284	1242	375	2.36	99.3
2012.08	470470	288	1374	357	2.61	99.4

2012.09	422046	294	903	337	1.76	99.6
2012.10	482101	282	752	361	1.45	99.6
2012.11	432681	361	756	510	1.40	99.2
2012.12	902928	859	1347	863	2.56	99.3
2013.01	433858	377	743	467	1.43	99.3
2013.02	476256	325	656	410	1.26	99.2
2013.03	444379	384	815	467	1.52	99.4
2013.04	479854	323	680	359	1.21	99.7
2013.05	422724	322	853	402	1.61	99.8
2013.06	728678	433	973	541	1.86	99.8
2013.07	569174	281	1104	315	2.04	100.0
2013.08	471411	298	1200	324	2.32	100.3
2013.09	431931	258	848	311	1.59	100.6
2013.10	482684	282	805	296	1.47	100.7
2013.11	436293	377	725	447	1.32	100.8
2013.12	905822	835	1351	933	2.62	100.9
2014.01	438646	431	703	530	1.38	100.7
2014.02	479268	365	612	446	1.21	100.7
2014.03	438145	397	891	476	1.66	101.0
2014.04	463964	304	630	401	1.15	103.1
2014.05	421117	348	846	432	1.56	103.5
2014.06	710375	356	933	394	1.72	103.4
2014.07	555276	304	1182	361	2.18	103.4
2014.08	463810	325	1159	356	2.12	103.7
2014.09	421809	319	840	383	1.51	103.9
2014.10	488273	345	707	391	1.33	103.6
2014.11	431543	398	716	519	1.33	103.2
2014.12	924911	892	1324	901	2.47	103.3
2015.01	440226	400	622	494	1.14	103.1

```
2015.02  488519  356  600  469 1.12 102.9
2015.03  449243  353  712  416 1.28 103.3
2015.04  476880  353  739  413 1.35 103.7
2015.05  430325  331  909  377 1.64 104.0
2015.06  733589  350  928  231 1.66 103.8
2015.07  587156  331 1105  347 2.02 103.7
2015.08  475369  339 1165  462 2.18 103.9
2015.09  415467  346  797  354 1.47 103.9
```

```
. gen t=_n                               <---- make data

. tsset t                                <---- set t as time series data
        time variable:  t, 1 to 189
                delta:  1 unit

. gen ry=y/(cpi/100)                     <---- real income

. gen rsp=(s/sml)/(cpi/100)              <---- real sake price per 1ml

. gen rbp=(b/bl)/(cpi/100)               <---- real beer price per 1l

. gen ds=0                               <---- dfault data

. replace ds=1 if f.rsp>rsp              <---- set ds=1 when excess demand exists
(94 real changes made)

. probit ds ry rsp rbp, if t<188.5 <---- Estimate probit
```

```
                                             during the period from 1 to 188

Iteration 0:   log likelihood = -130.30103
Iteration 1:   log likelihood = -95.883766
Iteration 2:   log likelihood = -95.419505
Iteration 3:   log likelihood = -95.419207
Iteration 4:   log likelihood = -95.419207

Probit regression                                 Number of obs   =        188
                                                  LR chi2(3)      =      69.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -95.419207                       Pseudo R2       =     0.2677

------------------------------------------------------------------------------
          ds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ry |   8.04e-07   9.42e-07     0.85   0.393    -1.04e-06    2.65e-06
         rsp |   -13.8574   2.166711    -6.40   0.000    -18.10408   -9.610729
         rbp |   .0026681   .0067234     0.40   0.691    -.0105094    .0158457
       _cons |    9.44494   3.697318     2.55   0.011     2.198331    16.69155
------------------------------------------------------------------------------
Note: 1 failure and 0 successes completely determined.

. logit ds ry rsp rbp if t<188.5 <---- Estimate logit
                                       during the period from 1 to 188

Iteration 0:   log likelihood = -130.30103
Iteration 1:   log likelihood = -96.132508
Iteration 2:   log likelihood =  -95.65503
```

```
Iteration 3:   log likelihood = -95.653538
Iteration 4:   log likelihood = -95.653538

Logistic regression                               Number of obs   =        188
                                                  LR chi2(3)      =      69.29
                                                  Prob > chi2     =     0.0000
Log likelihood = -95.653538                       Pseudo R2       =     0.2659

------------------------------------------------------------------------------
          ds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ry |   1.41e-06   1.56e-06     0.90   0.368    -1.65e-06    4.47e-06
         rsp |  -23.36485   3.941076    -5.93   0.000    -31.08922   -15.64048
         rbp |   .0046687   .0113128     0.41   0.680    -.0175041    .0268414
       _cons |   15.82621   6.301665     2.51   0.012     3.475172    28.17724
------------------------------------------------------------------------------
```

$D_t - S_t = \beta_0 + \beta_1 ry_t + \beta_2 rsp_t + \beta_3 rbp_t$

D_t is observed, but S_t is not observed. Therefore, $D_t - S_t$ is unobserved.

$rsp_{t+1} > rsp_t \implies D_t - S_t > 0 \implies ds_t = 1.$

$rsp_{t+1} \le rsp_t \implies D_t - S_t \le 0 \implies ds_t = 0.$

Example 2: Consider the two utility functions: $U_{1i} = X_i\beta_1 + \epsilon_{1i}$ and $U_{2i} = X_i\beta_2 + \epsilon_{2i}$.

A linear utility function is problematic, but we consider the linear function for simplicity of discussion.

We purchase a good when $U_{1i} > U_{2i}$ and do not purchase it when $U_{1i} < U_{2i}$.

We can observe $y_i = 1$ when we purchase the good, i.e., when $U_{1i} > U_{2i}$, and $y_i = 0$ otherwise.

$$
\begin{aligned}
P(y_i = 1) &= P(U_{1i} > U_{2i}) = P(X_i(\beta_1 - \beta_2) > -\epsilon_{1i} + \epsilon_{2i}) \\
&= P(-X_i\beta^* > \epsilon_i^*) = P(-X_i\beta^{**} > \epsilon_i^{**}) = 1 - F(-X_i\beta^{**}) = F(X_i\beta^{**})
\end{aligned}
$$

where $\beta^* = \beta_1 - \beta_2$, $\quad \epsilon_i^* = \epsilon_{1i} - \epsilon_{2i}$, $\quad \beta^{**} = \dfrac{\beta^*}{\sigma^*}$ and $\epsilon_i^{**} = \dfrac{\epsilon_i^*}{\sigma^*}$.

We can estimate β^{**}, but we cannot estimate ϵ_i^* and σ^*, separately.

Mean and variance of ϵ_i^{**} are normalized to be zero and one, respectively.

If the distribution of ϵ_i^{**} is symmetric, the last equality holds.

We can estimate β^{**} by MLE as in Example 1.

Example 3: Consider the questionnaire:

$$y_i = \begin{cases} 1, & \text{if the } i\text{th person answers YES,} \\ 0, & \text{if the } i\text{th person answers NO.} \end{cases}$$

Consider estimating the following linear regression model:

$$y_i = X_i\beta + u_i.$$

When $\mathrm{E}(u_i) = 0$, the expectation of y_i is given by:

$$\mathrm{E}(y_i) = X_i\beta.$$

Because of the linear function, $X_i\beta$ takes the value from $-\infty$ to ∞.

However, $\mathrm{E}(y_i)$ indicates the ratio of the people who answer YES out of all the people, because of $\mathrm{E}(y_i) = 1 \times P(y_i = 1) + 0 \times P(y_i = 0) = P(y_i = 1)$.

That is, $\mathrm{E}(y_i)$ has to be between zero and one.
Therefore, it is not appropriate that $\mathrm{E}(y_i)$ is approximated as $X_i\beta$.

The model is written as:

$$y_i = P(y_i = 1) + u_i,$$

where u_i is a discrete type of random variable, i.e., u_i takes $1 - P(y_i = 1)$ with probability $P(y_i = 1)$ and $-P(y_i = 1)$ with probability $1 - P(y_i = 1) = P(y_i = 0)$.

Consider that $P(y_i)$ is connected with the distribution function $F(X_i\beta)$ as follows:

$$P(y_i = 1) = F(X_i\beta),$$

where $F(\cdot)$ denotes a distribution function such as normal dist., logistic dist., and so on. $\longrightarrow$ probit model or logit model.

The probability function of y_i is:

$$f(y_i) = F(X_i\beta)^{y_i}(1 - F(X_i\beta))^{1-y_i} \equiv F_i^{y_i}(1 - F_i)^{1-y_i}, \qquad y_i = 0, 1.$$

The joint distribution of $y_1, y_2, \cdots, y_n$ is:

$$f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} F_i^{y_i}(1 - F_i)^{1-y_i} \equiv L(\beta),$$

which corresponds to the likelihood function. $\longrightarrow$ MLE

Example 4: Ordered probit or logit model:

Consider the regression model:

$$y_i^* = X_i\beta + u_i, \qquad u_i \sim (0, 1), \qquad i = 1, 2, \cdots, n,$$

where y_i^* is unobserved, but y_i is observed as $1, 2, \cdots, m$, i.e.,

$$y_i = \begin{cases} 1, & \text{if } -\infty < y_i^* \leq a_1, \\ 2, & \text{if } a_1 < y_i^* \leq a_2, \\ \vdots, & \\ m, & \text{if } a_{m-1} < y_i^* < \infty, \end{cases}$$

where a_1, a_2, $\cdots$, a_{m-1} are assumed to be known.

Consider the probability that y_i takes $1, 2, \cdots, m$, i.e.,

$$\begin{aligned}
P(y_i = 1) &= P(y_i^* \le a_1) = P(u_i \le a_1 - X_i\beta) \\
&= F(a_1 - X_i\beta), \\
P(y_i = 2) &= P(a_1 < y_i^* \le a_2) = P(a_1 - X_i\beta < u_i \le a_2 - X_i\beta) \\
&= F(a_2 - X_i\beta) - F(a_1 - X_i\beta), \\
P(y_i = 3) &= P(a_2 < y_i^* \le a_3) = P(a_2 - X_i\beta < u_i \le a_3 - X_i\beta) \\
&= F(a_3 - X_i\beta) - F(a_2 - X_i\beta), \\
&\vdots \\
P(y_i = m) &= P(a_{m-1} < y_i^*) = P(a_{m-1} - X_i\beta < u_i) \\
&= 1 - F(a_{m-1} - X_i\beta).
\end{aligned}$$

Define the following indicator functions:

$$I_{i1} = \begin{cases} 1, & \text{if } y_i = 1, \\ 0, & \text{otherwise.} \end{cases} \quad I_{i2} = \begin{cases} 1, & \text{if } y_i = 2, \\ 0, & \text{otherwise.} \end{cases} \quad \cdots \quad I_{im} = \begin{cases} 1, & \text{if } y_i = m, \\ 0, & \text{otherwise.} \end{cases}$$

More compactly,

$$P(y_i = j) = F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta),$$

for $j = 1, 2, \cdots, m$, where $a_0 = -\infty$ and $a_m = \infty$.

$$I_{ij} = \begin{cases} 1, & \text{if } y_i = j, \\ 0, & \text{otherwise,} \end{cases}$$

for $j = 1, 2, \cdots, m$.

Then, the likelihood function is:

$$L(\beta) = \prod_{i=1}^{n} \Big(F(a_1 - X_i\beta)\Big)^{I_{i1}} \Big(F(a_2 - X_i\beta) - F(a_1 - X_i\beta)\Big)^{I_{i2}} \cdots \Big(1 - F(a_{m-1} - X_i\beta)\Big)^{I_{im}}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \Big(F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)\Big)^{I_{ij}},$$

where $a_0 = -\infty$ and $a_m = \infty$. Remember that $F(-\infty) = 0$ and $F(\infty) = 1$.

The log-likelihood function is:

$$\log L(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} \log\Big(F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)\Big).$$

The first derivative of $\log L(\beta)$ with respect to β is:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{-I_{ij} X_i' \Big(f(a_j - X_i\beta) - f(a_{j-1} - X_i\beta)\Big)}{F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)} = 0.$$

Usually, normal distribution or logistic distribution is chosen for $F(\cdot)$.

Example 5: Multinomial logit model:

The ith individual has $m+1$ choices, i.e., $j = 0, 1, \cdots, m$.

$$P(y_i = j) = \frac{\exp(X_i\beta_j)}{\sum_{j=0}^{m} \exp(X_i\beta_j)} \equiv P_{ij},$$

for $\beta_0 = 0$. The case of $m = 1$ corresponds to the bivariate logit model (binary choice).

Note that

$$\log \frac{P_{ij}}{P_{i0}} = X_i\beta_j$$

The log-likelihood function is:

$$\log L(\beta_1, \cdots, \beta_m) = \sum_{i=1}^{n} \sum_{j=0}^{m} d_{ij} \ln P_{ij},$$

where $d_{ij} = 1$ when the ith individual chooses jth choice, and $d_{ij} = 0$ otherwise.

Example 6: Nested logit model:

(i) In the 1st step, choose YES or NO. Each probability is P_Y and $P_N = 1 - P_Y$.

(ii) Stop if NO is chosen in the 1st step. Go to the next if YES is chosen in the 1st step.

(iii) In the 2nd step, choose A or B if YES is chosen in the 1st step. Each probability is $P_{A|Y}$ and $P_{B|Y}$.

For simplicity, usually we assume the logistic distribution.

So, we call the nested logit model.

The probability that the ith individual chooses NO is:

$$P_{N,i} = \frac{1}{1 + \exp(X_i\beta)}.$$

The probability that the ith individual chooses YES and A is:

$$P_{A|Y,i}P_{Y,i} = P_{A|Y,i}(1 - P_{N,i}) = \frac{\exp(Z_i\alpha)}{1 + \exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

The probability that the ith individual chooses YES and B is:

$$P_{B|Y,i}P_{Y,i} = (1 - P_{A|Y,i})(1 - P_{N,i}) = \frac{1}{1 + \exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

In the 1st step, decide if the ith individual buys a car or not.
In the 2nd step, choose A or B.

X_i includes annual income, distance from the nearest station, and so on.
Z_i are speed, fuel-efficiency, car company, color, and so on.

The likelihood function is:

$$L(\alpha,\beta) = \prod_{i=1}^{n} P_{N,i}^{I_{1i}} \Big(((1-P_{N,i})P_{A|Y,i})^{I_{2i}} ((1-P_{N,i})(1-P_{A|Y,i}))^{1-I_{2i}} \Big)^{1-I_{1i}}$$
$$= \prod_{i=1}^{n} P_{N,i}^{I_{1i}} (1-P_{N,i})^{1-I_{1i}} \Big(P_{A|Y,i}^{I_{2i}} (1-P_{A|Y,i})^{1-I_{2i}} \Big)^{1-I_{1i}},$$

where

$$I_{1i} = \begin{cases} 1, & \text{if the } i\text{th individual decides not to buy a car in the 1st step,} \\ 0, & \text{if the } i\text{th individual decides to buy a car in the 1st step,} \end{cases}$$
$$I_{2i} = \begin{cases} 1, & \text{if the } i\text{th individual chooses A in the 2nd step,} \\ 0, & \text{if the } i\text{th individual chooses B in the 2nd step,} \end{cases}$$

Remember that $\mathrm{E}(y_i) = F(X_i\beta^*)$, where $\beta^* = \dfrac{\beta}{\sigma}$.
Therefore, size of β^* does not mean anything.

The marginal effect is given by:

$$\frac{\partial \mathrm{E}(y_i)}{\partial X_i} = f(X_i\beta^*)\beta^*.$$

Thus, the marginal effect depends on the height of the density function $f(X_i\beta^*)$.