Lecture 6: statistics

Quantitative measurement and analysis of actual phenomena:

Describe a phenomenon (Specification of a relationship between a dependent and a series of independent variables)

Test a hypothesis (Evaluation of alternative theories with quantitative evidence)

Forecasting (Prediction under different scenarios

Modelling process – Steps:

1 Theoretical background - Hypotheses development

2 Statistical model (Model specification): The functional form of the model [y = f (x)].

3 Data / data collection

4 Model estimation

5 Hypothesis testing

6 Model interpretation

Application examples:

Example 1: Trip generation

A trip generation model can:

– Provide insights about the most important factors that affect trip generation

– Provide predictions for trip generation per household

Transportation planners can make more informed decisions about infrastructure investments, traffic management

Example 2: Mode choice

Mode choice models can:

– Provide insights about the most influential factors in mode choice

– Forecast demand for current and future solutions

– Can inform the design and evaluation of transportation policies

– Help to evaluate the effectiveness of existing transportation policies

Example 3: Road safety

Car crash occurrence and severity models are essential for improving road safety

Understand the factors that contribute to car crashes (e.g. driver behavior)

Can inform the design and evaluation of road safety interventions, such as improving road infrastructure

Example 4: Policy making in railways

Railway econometrics is essential for understanding the economic aspects of railway transportation

Can support efficient and sustainable railway operations

Can inform the development of new business models and revenue streams for railways

Example 5: Driving behavior:

We model longitudinal an latitudinal behavior of drivers at the individual level

Typical examples:

Car-following (acceleration decisions with respect to the behaviour of a lead vehicle)

Lane-changing (decision making process with respect to lane-changing behaviour)

Gap acceptance (driver behaviour at intersections, roundabouts etc.)

Regression analysis:

Statistical technique to "explain" movements (changes) in a variable (dependent variable)

$Yi = \beta 0 + \beta 1Xi + \varepsilon i$

Where εi is independent and identically normally distributed

Regression assumptions:

Assumption I: The dependent variable is linear in parameters β:

Model interpretation: Change of one unit in the independent variable results in β change to the dependent variable, all others being equal. The linear effect of independent variables only holds for linear regression models.

Assumption II: The error term εi has a zero mean: $Yi = \beta 0+\beta 1Xi + \varepsilon i$, $E(Yi|Xi)=\beta 0+\beta 1Xi$, $E(\varepsilon i|Xi)=0$

Interpretation: The mean of ε does not have any effect on Y

Assumption III: Homoskedasticity

Variance of the error term var(ε) = σ2 is constant across observation and uncorrelated to the independent variables

Assumption IV: No autocorrelation of the error terms:

The error terms across different observations are not correlated: $cov(\varepsilon\ \varepsilon) = 0\ \forall i \neq j$

Implication: There is no correlation between the Ys, cov(Yi Yj) = 0

Assumption V: No correlation between the error term and the independent variables:

cov(εi Xi) = 0 for all X variables

Assumption VI: The error terms are approximately normally distributed: The distribution of the error terms is ε~N(0,σ2)

Summary of regression assumptions:

Statistical Assumption	**Mathematical Expression**
1. Functional form	$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$
2. Zero mean of disturbances	$E[\varepsilon_i] = 0$
3. Homoscedasticity of disturbances	$\mathrm{VAR}[\varepsilon_i] = \sigma^2$
4. Nonautocorrelation of disturbances	$COV[\varepsilon_i, \varepsilon_j] = 0$ if $i \neq j$
5. Uncorrelatedness of regressor and disturbances	$COV[X_{1i}, \varepsilon_j] = 0$ for all I and j
6. Normality of disturbances	$\varepsilon_i = N(0, \sigma^2)$

A short note on dummy variables:

the independent variable may not be continuous but categorical for instance gender, education level….. Dummy variable: We can estimate one less parameter than the total number of groups of the independent variable (n-1) to avoid perfect multicollinearity

Fundamentals of logit models:

Dependent variable: A discrete outcome that indicates the presence of a condition

Outcome: Population proportion or probability (P) for the occurrence of the selected outcome

The concept of utility theory is used in many transportation applications:

– Alternatives (conditions) have a utility based on attributes (e.g. travel time or travel cost)

– Utility affects behaviour (choice or presence of condition)

Maximum likelihood estimation (MLE)

The LL function as a function of β is specified as: $LL(\beta) = \sum_{n=1}^{N} ln(P_{n,j_n}(\beta))$

MLE example for linear regression

Let's consider the following linear regression model $Yi = \beta 0+\beta 1X1i + \beta 2X2i + \varepsilon i$,
with $Y \sim N(\mu,\sigma 2)$.

The PDF for normally distributed variables is $f(Y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5(\frac{Y-\mu}{\sigma})^2}$

For N observations, the linear regression model is:

Y1 = β0 + β1X11 + β2X21 + ε1

Y2 = β0 + β1X12 + β2X22 + ε2 .

……

YN =β0+β1X1N +β2X2N +εN

After replacing the regression model in the density function we get:

$$f(Y_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5(\frac{Y_1-(\beta_0+\beta_1 X_{11}+\beta_2 X_{21})}{\sigma})^2}$$

$$f(Y_2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5(\frac{Y_2-(\beta_0+\beta_1 X_{12}+\beta_2 X_{22})}{\sigma})^2}$$

$$\vdots$$

$$f(Y_N) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5(\frac{Y_N-(\beta_0+\beta_1 X_{1N}+\beta_2 X_{2N})}{\sigma})^2}$$

The joint probability (density) of all observations: $L(\beta) = \prod_{i=1}^{N}(f_i(\beta))$

maximise the logarithm ln, hence: $LL(\beta) = \sum_{i=1}^{N} ln(f_i(\beta))$

Confidence intervals:

Central limit theorem: When a sufficiently large random sample is drawn from a population with mean μ and standard deviation σ, the sample mean X is approximately normally distributed with mean μ and standard deviation σ/ √n. $0.95 = P\left(-1.96 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

Cumulative Density Function: The CDF of a probability distribution contains the probabilities that a random variable X is less than or equal to a given value x $F_X(x) = P[X \le x] = \int_{-\infty}^{x} f_X(x)\,dx$, for all $x \in \mathbb{R}$

Basic goodness-of-fit statistics:

ρ2 measure (bounded between 0 and 1) $\rho^2 = 1 - \frac{LL(\hat{\theta})}{LL(0)}$ if ρ2 = 1 then perfect fit.

Not to be mixed with the linear regression coefficient of determination R^2 which is also between 0 and 1. $R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}} = 1 - \frac{\sum(y_i - fi)^2}{\sum(y_i - \bar{y})^2}$

The AIC and BIC tests:

AIC=2k−2ln(L), k is the number of parameters in the model, L is likelihood, Lower AIC, better

BIC=ln(n)k−2ln(L), n is the number of observations or sample size, lower BIC, better

Lecture 7:

Driving behaviour models:

Acceleration-deceleration (i.e. car-following and free speed)

Lane-changing models (merging is a special category of lane-changing models)

Gap-acceptance models (e.g. junction crossing or roundabout merging)

Driving behaviour models		
Acceleration models	Gap-acceptance models	Lane-changing models
Car-following models	Intersection	Basic lane-changing manoeuvre
Free-flow models	Roundabout	Merging

Acceleration - car-following models:

Dependent variable: Acceleration

Type of dependent variable: Continuous

Modelling approach: Regression model (linear, non-linear)

Typical explanatory variables: Subject's speed, relative speed with lead vehicle, distance to lead vehicle, time-distance to lead vehicle, traffic density

Gap-acceptance models

Dependent variable: Gap-acceptance

Type of dependent variable: – Discrete (accept/reject a gap),

– Continuous (accept/reject a gap of a given size

Modelling approach: Logit model for discrete or regression-type model for continuous (CDF instead of PDF)

Typical explanatory variables: Gap size (for discrete treatment), distance of conflicting vehicle, speed of conflicting vehicle, acceleration of conflicting vehicle, type of conflicting vehicle

Lane-changing models

Dependent variable: Lane-changing manoeuvre

Type of dependent variable: Discrete-continuous

Modelling approach: Combined logit-regression models

– A driver decides a target lane (lane choice)

– A driver decides whether to accept the available gap to perform a lane-changing manoeuvre

Typical explanatory variables: Average lane density, traffic speed, specific lane characteristics, path-plan, presence and speed of adjacent vehicles

Lecture 8:

Linear regression model of trip generation - Dependent variable: Number of trips per household

Multinomial logit model of mode choice - Dependent variable: Choice between bus, metro and tram

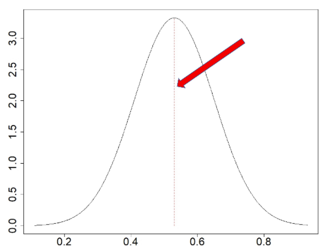
Random effects:

The concept of random effects is most commonly used to capture unobserved heterogeneity (heterogeneity in the dependent variable that is not explained or partially explained by the independent variables).

In theory, the simplest implementation would be to estimate one coefficient per person,

An alternative implementation is through the use of distributions

Else, what if instead of estimating a separate coefficient for each individual, we assumed that all individual coefficients follow some distribution with moments



Mixture probability distribution function: Convex combination of other probability distribution functions.

Mathematics of random effects

Discrete mixtures

Let's assume wi and i = 1, ..., N are positive weights so that: ∑wi =1 $g(\epsilon;\theta_1,\dots,\theta_n) = \sum_{i=1}^{N} w_i f(\epsilon,\theta_i)$

Continuous mixtures

Let's assume w(θ) a positive function so that: ∫w(θ)dθ=1 $g(\varepsilon)=\int_{\theta} w(\theta) f(\varepsilon,\theta)\,d\theta$

Linear model with continuous dependent variable example

A typical regression model takes the form: Yni =β0+β1X1ni +β1X2ni +εni

Now let's assume that we are adding an additional error component :

Yni =β0+β1X1ni +β1X2ni +ξn+εni, where ξn is an individual specific normally distributed error with zero mean as ξn =N(0,σ2ξ) and ε is i.i.d. normally distributed as εni =N(0,σ2ε)

εni is i.i.d. normally distributed. The distribution is random across individuals and observations. ξn is normally distributed across individuals i.e. for all observations of the **same individual, ξn takes the same** numerical value.

In a linear model with continuous dependent variable, ξn would shift the observations of an individual (plus or minus) by ξn. The variance of these shifts is σ2ε .

In practice, we could say that in a random intercept model, intercepts (model constants) are distributed as N(β0,σ2ε)

The same principle also applies in logit models

U1n = βX1,n + ξ1n + ε1n. U2n = βX2,n + ξ2n + ε2n U3n = βX3,n + ξ3n + ε3n

$$P(1|\xi X)=\frac{e^{\beta X_{1,n}+\xi_{1n}}}{e^{\beta X_{1,n}+\xi_{1n}}+e^{\beta X_{2,n}+\xi_{2n}}+e^{\beta X_{3,n}+\xi_{3n}}}$$

Let's assume: Yn,i =β0+β1X1n,i +εni =β0+(β1+ξ1)X1n,i +εni

Monte Carlo integration:

If a random variable x follows a distribution and has a density function f (x), then for another function g(x) we have: $\int_x g(x)f(x)\,dx=\frac{1}{R}\sum_{r=1}^{R} g(x_r)$

The likelihood function is: $L(\xi)=\prod_{n=1}^{N}\int_{\xi} f(\beta,\xi)\phi(\xi)\,d\xi=\frac{1}{R}\sum_{r=1}^{R}\prod_{n=1}^{N} f(\beta\xi^r)$

In many cases we have multiple observations per individual (panel data).

– In cross-sectional data (one observation per individual), we attempt to maximise the contribution of the likelihood of each observation

– In panel data, we want to maximize the contribution of the observations of one individual in the LL function.

The log-likelihood function in panel data is: $LL=\sum_{n=1}^{N}\left(\sum_{t=1}^{T} f_{n,t}(\beta)\right)$

Ynt observations where N is the number of individuals and T the number of responses per individual.

$$L(\xi)=\prod_{n=1}^{N}\int_{\xi}\left[\prod_{t=1}^{T} f(\beta,\xi)\phi(\xi)\right]d\xi$$

To approximate an integral with Monte Carlo simulation, we must:

=>generate R draws of the target distribution

=>average probability values across draws per individual

A normally distributed random effect (parameter) is specified: βrandom = betaμ + betaσ ∗ draw^r

One can easily generate log-normally distributed random draws by taking the exponential of normal draws as: $e^{beta_{\mu}+beta_{\sigma}*draw^{r}}$

Truncation: Sometimes, we want to simulate numbers of a distribution from a specific range (a, b). A generic formula to do this is: $r = \Phi^{-1}(\Phi(a|\beta) + U * (\Phi(b|\beta) - \Phi(a|\beta))|\beta)$

Application 1: Taste heterogeneity

Application 2: Correlated observations

If we assume we have a series of correlated error terms ε_{ni}, we can then:

$$Y_{ni} = \beta_0 + \beta_1 X_{1ni} + \beta_1 X_{2ni} + \varepsilon_{ni} = \beta_0 + \beta_1 X_{1ni} + \beta_1 X_{2ni} + \xi_n + \varepsilon_{ni}$$

On top of the random error ε, there is an error term ξ which always takes the same value for an individual but varies across individuals following a distribution.

In the model specification, we typically consider $\xi = \alpha * draw_r$ where α is a parameter to be estimated, and $draw_r$ is the rth draw of R

Let's assume the utility for accepting a gap is $U_{nt} = V_{nt} + \varepsilon_{nt} = \beta X_{nt} + \xi_{nt} + \varepsilon_{nt}$

Let's revisit the formula of the GM model: $\alpha_n(t) = \alpha \frac{V_n(t)^\beta}{\Delta X_n(t)^\gamma} |\Delta V_n(t - \tau_n)|^\lambda$

Truncation is required in the density function of reaction time $f(\tau_n) = \begin{cases} \dfrac{\frac{1}{\tau_n \sigma_\tau} \phi\left(\frac{ln(\tau_n) - \mu_\tau}{\sigma_\tau}\right)}{\Phi\left(\frac{ln(\tau^{max}) - \mu_\tau}{\sigma_\tau}\right) - \Phi\left(\frac{ln(\tau^{min}) - \mu_\tau}{\sigma_\tau}\right)} & \text{if } \tau^{min} < \tau_n \leq \tau^{max} \\ 0 & otherwise \end{cases}$

Reaction time as random effect:

Reaction time, suggests how much backwards we must go in the data and get a value for an attribute - the attribute value changes depending on the reaction time.

Example: If $\tau = 1.27$ then I must go 1.27 backwards in my data

Specification process (1)

- Define independent and dependent variables
- Generate R reaction time values for each individual
- Go backwards R times for each individual and take the "lagged" value
- If it is not possible to go backwards by τ exactly:
    - For example if $\tau = 1.27$ and I have 1 observations per second
    - If $\tau = 1.27$ then I will extract the values 1s and 2s before the current timestamp and I will calculate for 1.27 using some formulae from kinematics

Specification process (2)

- Now that I have extracted my "lagged" variables, I will compute the model density R times, for each τ_r that I generated
- Then we must take the product of each density function observation per individual n and per draw r
- Last step: solve the integral

Some notes on integration:

The average across draws to approximate the integral applies ONLY for Monte Carlo simulation

That is because the draws are generated already following some distribution and hence, are weighted

Simulation may not be the most suitable technique for reaction time:

1. With simulation we update the distribution in each iteration of estimation.
2. if we update the reaction time distribution (hence draws) each iteration then we must go "backwards" in the data for different time intervals (potentially very time consuming)
3. Instability: Based on the estimated μ and σ we may generate a very narrow distribution of τ that does not allow us to go sufficiently backwards.

Numerical integration:

A generic rule of numerical integration techniques is $\int f(x)\,dx \approx \sum_{i=1}^{N} w_i f(x_i)$

Gauss-Legendre quadrature

$\int_{-1}^{1} f(x)\,dx \approx \sum_{i=1}^{N} w_i f(x_i)$ The formula is extended to any range as: $\int_{a}^{b} f(x)\,dx \approx \frac{b-a}{2} \sum_{i=1}^{N} w_i f(\frac{b-a}{2}x_i + \frac{b+a}{2})$

1. Model specification with Gauss-Legendre quadrature
2. We generate nodes xi at the (-1, 1) range and then transform them to the (a, b),
3. We use these reaction time values to go "backwards"
4. We follow the specification process until the step of multiplying probability (or density) values per individual per draw as before.
5. We write the density function of reaction time in the model specification (truncated log-normal). or each individual, we multiply the product of observations per draw with the reaction time density function value
6. We compute the weighted average
7. We compute the log-likelihood

Is there a solution in-between?

Simulation is convenient for implementation and easier to add additional integral dimensions, However, instabilities and slow

Numerical integration is faster however suffers from the curse of dimensionality, there can be estimation issues due to the incorporation of τ pdf.

Importance sampling:

The idea is that instead of a target distribution f () that is difficult to draw from, we draw from a proposal distribution g()

We draw R draws from g() and follow the same process as Monte Carlo simulation by weighting our likelihood function as f(xr)/g(xr)

Requirements:

The f () and g () distributions have the same support (min and max range)

The ratio f()/g() is always finite for every draw xr

Importance sampling in car-following models: steps:

We decide on a fixed distribution to draw from

The moments of the distribution is irrelevant as long as it covers the range of reaction time

We use these draws before estimation

We add the f ()/g () in the model specification

看下代码吧，还是不太懂

Lecture 9

Latent class models - discrete mixtures

Mathematics of random effects:

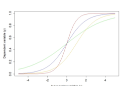
Continuous mixtures: assume w(θ) a positive function so that: $\int_{\theta} w(\theta)\,d\theta = 1$ $g(\epsilon) = \int_{\theta} w(\theta) f(\epsilon,\theta)\,d\theta$

Discrete mixtures: Let's assume wi and i = 1, ..., N are positive weights so that: $\sum_{i=1}^{N} w_i = 1$ $g(\epsilon;\theta_1,\ldots,\theta_n) = \sum_{i=1}^{N} w_i f(\epsilon,\theta_i)$

Let's assume a binary choice model of choosing alternative i: $P_i(\beta_1) = \frac{1}{1+e^{(-X\beta_1)}}$

$P_i(\beta_1) = \frac{1}{1+e^{(-X\beta_1)}}$ $P_i(\beta_2) = \frac{1}{1+e^{(-X\beta_2)}}$

$P_i(\beta_3) = \frac{1}{1+e^{(-X\beta_3)}}$ $P_i(\beta_4) = \frac{1}{1+e^{(-X\beta_4)}}$



– Let's assume we want: 0.3P(β1)+0.3P(β2)+0.2P(β3)+0.2P(β4)

– We need wi and i = 1, ..., N are positive weights so that: $\sum wi=1$

Latent classes:

– The logit formula can be used to represent the weights wi

– Let's assume that the sample can be grouped in S groups (classes), each class with a model that best captures observed patterns

– We can define a likelihood function such as. $$L_n(\beta\pi) = \sum_{s=1}^{S} \pi_{ns} \left(\prod_{t=1}^{T_n} P_{ni^*t}(\beta_s) \right)$$

– The class membership probability π is estimated as a parameter

– We can use the logit formula to constrain the π values to positive

– We can also independent variables to further explain class membership probabilities

$$\pi_{ns} = \frac{e^{\delta_s + g(\gamma_s, z_n)}}{\sum_{l=1}^{S} e^{\delta_l + g(\gamma_l, z_n)}}$$

Latent class example - 2 classes

	Class 1		Class 2		Class allocation		
	Estimate	t-ratio	Estimate	t-ratio		Estimate	t-ratio
asc_{bus}	0.12	5.35	-0.08	-3.40	Income	-0.03	-1.77
asc_{train}	0.22	2.11	-0.15	-7.31	Age	-0.01	-1.97
tt_{car}	-0.03	-1.86	-0.02	-7.93	Car ownership	-0.32	-2.35
tc_{car}	-0.05	-6.37	-0.01	-8.40			
tt_{bus}	-0.06	-2.12	-0.09	-0.23			
tc_{bus}	-0.03	-3.93	-0.02	-3.81			
tt_{train}	-0.05	-1.96	-0.06	-1.74			
tc_{train}	-0.06	-1.07	-0.01	-1.77			

Latent classes summary

Latent class and car-following

How can we specify a truncated distribution?

$$f(x)^* = \frac{f(x)}{F(b)-F(a)}$$

where:

$f()$ is the density function of variable x

$F()$ is the cumulative distribution function of variable x

a, b define the minimum and maximum values of truncation

We can determine the probability of each state using the logit model

What are the truncation bounds for acceleration and deceleration?

– Truncated normal distribution for acceleration:

$$\phi(x^{acc})^* = \frac{\frac{1}{\sigma^{acc}} \phi\left(\frac{x^{acc}-\mu^{acc}}{\sigma^{acc}}\right)}{1-\Phi\left(\frac{-\mu^{acc}}{\sigma^{acc}}\right)}$$

– Truncated normal distribution for deceleration:

$$\phi(x^{dec})^* = \frac{\frac{1}{\sigma^{dec}} \phi\left(\frac{x^{dec}-\mu^{dec}}{\sigma^{dec}}\right)}{\Phi\left(\frac{-\mu^{dec}}{\sigma^{dec}}\right)}$$

The total probability of an observation for an individual n at time t is:

P =Paccφ(xacc)∗+Pdecφ(xdec)∗+Pdnφ(xdn)

However, we need to assign probabilities to the acceleration-deceleration observations in the data

P = [Paccφ(xacc)∗+Pdnφ(xdn)∗](Acceleration≥0)

+ [P dec φ(x dec)∗ + P dn φ(x dn)∗](Acceleration < 0)

$$P_{nt} = [P_{nt}^{acc} \phi(x_{nt}^{acc})^* + P_{nt}^{dn} \phi(x_{nt}^{dn})^*](Acceleration \geq 0) \\ + [P_{nt}^{dec} \phi(x_{nt}^{dec})^* + P_{nt}^{dn} \phi(x_{nt}^{dn})^*](Acceleration < 0)$$

Model transferability: The transfer of a model estimated in one context to a different one.

Motivation: To reduce the efforts in the model development (using the same structure)

To reduce or eliminate the need for a large data collection in the application context.

Transferability evaluation - Parameter equivalence

t-test of individual parameter equivalence

$$t_{diff} = \frac{\beta_k - \beta_{k*}}{\sqrt{\sigma_k^2 + \sigma_{k*}^2}}$$

βk: parameter estimates of the estimation (transferred) context model

βk∗: parameter estimates of the application context model

σk and σk∗, standard errors

If |tdiff | > 1.96 then significant difference

Transferability evaluation - Transferability Test Statistic (TTS)

assesses whether the null hypothesis of statistical equivalence between the transferred and the application context model is rejected or not: TTS =−2[LLk∗(βk)−LLk(βk)]

LLk∗: log-likelihood on the application context data using the transferred context parameters.

LLk: log-likelihood on the application context data using application context parameters

Bayesian updating:

$$\beta_{upt} = \left(\frac{\beta_k}{\sigma_k^2} + \frac{\beta_{k*}}{\sigma_{k*}^2}\right)\left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_{k*}^2}\right)^{-1}$$

Combined transfer estimation:

$$\beta_{upt} = \left(\frac{\beta_k}{\sigma_k^2 + \alpha\alpha'} + \frac{\beta_{k*}}{\sigma_{k*}^2}\right)\left(\frac{1}{\sigma_k^2 + \alpha\alpha'} + \frac{1}{\sigma_{k*}^2}\right)^{-1}$$

where: $\alpha = \beta_k - \beta_{k*}$ and $\alpha' = \beta_{k*} - \beta_k$

Joint model estimation

Motivation: The joint estimation of models using various data sources

The datasets can have different variables e.g. one of the model variables may be unavailable in one of the data sets.

The overall process is:

For logit model we estimate different constants and scale parameters for each data set. We keep the same parameters for the independent variables.

For continuous models we estimate different constants and standard deviation parameters.

For continuous model and data sets a and b we have:

Ya = β0a + β1X1a + β2X2a

Yb = β0b + β1X1b + β2X2b and

$$f(Y_a) = \frac{1}{\sigma_a}\phi\left(\frac{Y_a - \widehat{Y}_a}{\sigma_a}\right)$$

$$f(Y_b) = \frac{1}{\sigma_b}\phi\left(\frac{Y_b - \widehat{Y}_b}{\sigma_b}\right)$$

$$f(Y) = (Y_a)^{(data==a)}(Y_b)^{(data==b)}$$

model evaluation:

1. We estimate the joint model as described before (k parameters in total).
2. We estimate the model on Ya only (n parameters in total).
3. We estimate the model on Yb only (n parameters in total).
4. we compute the combined LL from steps 2 and 3 (LLY +LLY)
5. we take the LL value from step 1 as LLYab
6. We calculate the LR test as $-2[(LL_{Y_a} + LL_{Y_b}) - LL_{Y_{ab}}]$ and evaluate for degrees of freedom 2nk.

Lecture 10

Goodness-of-fit:

Likelihood ratio (LR) test: LR = −2(LLrestricted − LLunrestricted)
– For instance, I estimated a model and got LLrestricted = −6434.891
– Then, I added 3 new independent variables and received LLunrestricted = −6177.035
LR = −2(LLrestricted − LLunrestricted) = −2(−6434.891 − [−6177.035]) = 515.712
We must evaluate this result for 3 degrees of freedom (DoF are equal to the number of newly added parameters) at the 0.05 level of significance
– We find the critical values in the χ2-distribution table If LR > χ2 critical value then the addition of the new variable is significantly improving model fit, 515.712 > 7.81 hence, significantly improve
Akaike Information Criterion (AIC): AIC=−2ln(L)+2k, Bayesian (BIC): BIC=−2ln(L)+kln(n)

t-test of individual parameter equivalence: $t_{diff} = \frac{\beta_k - \beta_{k*}}{\sqrt{\sigma_k^2 + \sigma_{k*}^2}}$

βk: parameter estimates of model k
βk∗: parameter estimates of model k*
σk and σk∗, standard errors
If |tdiff | > 1.96 then significant difference

Comparison of models overall:
– The test can be always used for the continuous models that we have seen so far.
– The test of individual parameters cannot be used in logit models
– The logit model parameters include a scale coefficient μ which cannot be estimated and differs across models
– Alternatively evaluate model overall:

We have Model 1 and Model 2 and want to see how close they are
Model 1 is estimated from dataset1 and Model 2 is estimated from dataset 2
We use the results of Model 2 as starting values on dataset1 (or the opposite)
We take note of the initial log-likelihood value
We implement the LR test with degrees of freedom equal to the number of parameters (of Model 2)
The LLunrestricted is the LL of Model 1
The LLrestricted is the initial LL when the results of Model 2 are implemented as starting values on dataset 1

Lane-changing models:
Two types of lane-changing models

Modelling lane-changing decision making process
Modelling the impact on the surrounding traffic

Modelling lane-changing decision making process
Rule based (Gipps'-type) models: Risk of collision, Obstructions Heavy vehicles

Special purpose lanes, Intention for taking an exit, Speed advantage

Yang and Koutsopoulos (1996), probabilistic approach:

Mandatory lane-changing: (不得不变道 为了转弯之类的)

A probability for a mandatory lane-change to take place
A gap-acceptance model based on deterministic critical gap rules

Discretionary lane-changing

Whether lane-changing will improve current travel attributes (e.g allow for higher speed)
A gap-acceptance model is also considered

Utility-based models

– Ahmed (1999) developed a more flexible version of the lane-changing model of Yang and Koutsopoulos (1996)

– Toledo (2003), merged the discretionary and mandatory lane-changing models in the same specification

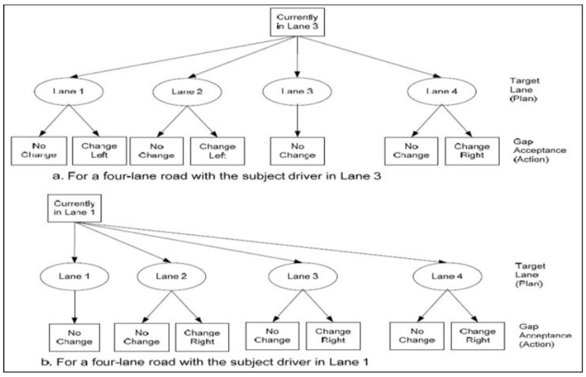
– Choudhury (2007) introduced the concept of latent plans

Latent plans lane-changing

– We observe the actions of a driver (lane-change or not) but not the intention Examples:

A driver executes a lane-changing manoeuvre aiming to reach an even further lane

A driver stays in a lane not because it is the target lane but because traffic conditions do not allow for a lane-changing manoeuvre



Model components:

Target lane model:

The factors affecting target lane utility can be:

– Lane attributes

– Neighbouring traffic attributes

– Path-plan attributes

A driver chooses the lane with the highest utility as the target lane The typical linear utility function Unt = Vnt + εnt can be used to determine utility. $P_n(l_t) = \frac{e^{V_{lnt}}}{\sum_{l' \in L_n} e^{V_{l'nt}}}$

Gap-acceptance model:

The direction of lane-change is determined by the lane with the higher utility

Then, a driver evaluates whether the traffic conditions allow for a lane-changing manoeuvre towards that direction

Critical gap: The minimum gap size a driver would accept to perform a lane-changing

Critical gap is not observed - approximated via explanatory variables

Usually assumed to follow a log-normal distribution

Gap-acceptance model: $P_n[G^g_{lnt} > G^{gcr}_{lnt}] = P_n[ln(G^g_{lnt}) > ln(G^{gcr}_{lnt})] = \Phi\left[\frac{ln(G^g_{lnt}) - \beta X^g_{lnt}}{\sigma_g}\right]$

$$P^{accept}_n = P_n[ln(G^{lead}_{lnt}) > ln(G^{cr,lead}_{lnt})]P_n[ln(G^{lag}_{lnt}) > ln(G^{cr,lag}_{lnt})] = \Phi\left[\frac{ln(G^{lead}_{lnt}) - \beta X^{lead}_{lnt}}{\sigma_{lead}}\right]\Phi\left[\frac{ln(G^{lag}_{lnt}) - \beta X^{lag}_{lnt}}{\sigma_{lag}}\right]$$

The probability of rejecting any of the adjacent gaps is Preject =1−Paccept

The total probability of a lane-changing manoeuvre is: $P^{total}_n = P_n(l_t)P^g_n \;, g \in (lead, lag)$

where: $P_n(l_t) = \sum_{l \in L_n} P_n$ towards the direction of the target lane