LEC003 Demand Forecasting

linear reg

ARIMA

Bass Diffusion

Multinomial logit

Boosting

XGBoost

VG441 SS2020

Cong Shi
Industrial & Operations Engineering
University of Michigan

Linear Regression

- "Best fitting line"

$y = h_\theta(x) = \theta_0 + \theta_1 X$ *

其中$\theta = (\theta_0, \theta_1)$

X可以为$(X_1, \cdots, X_n)$

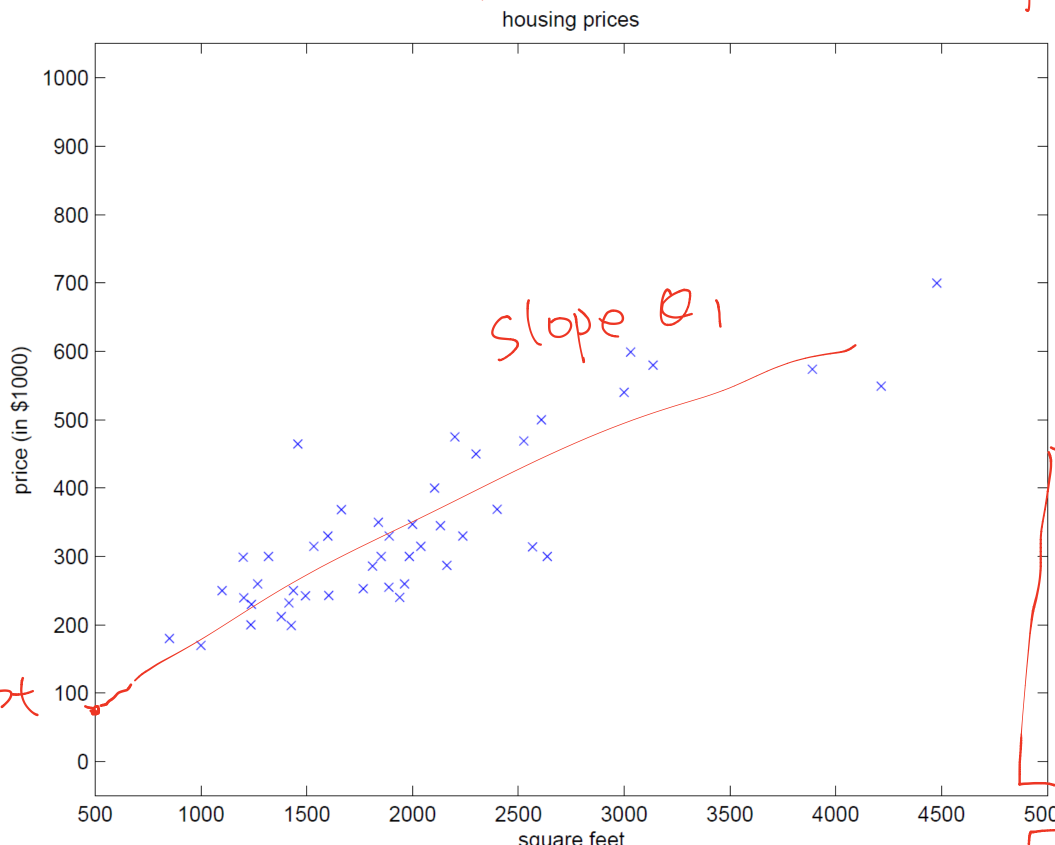
y

response

dep var

regressand

hypothesis function



intercept θ_0

explanatory var.

indep var

regressor

features

covariate

X

Linear Regression

$X = (x_1, \cdots x_n)$ n维

房地产 e.g. 面积 location

- (Linear) Hypothesis Function:

$$h_\theta(x) = \theta_0 \underbrace{x_0}_{} + \theta_1 x_1 + \cdots + \theta_n x_n \qquad x_0 = 1$$

- Minimizing the least-squares cost: 求导

$$J\left(\theta_{0\ldots n}\right) = \frac{1}{2m} \sum_{i=1}^{m} \left(\underbrace{h_\theta\left(x^{(i)}\right)}_{\text{line}} - \underbrace{y^{(i)}}_{\text{real value}} \right)^2$$

随m组 X, y.

Matrix Derivation

- Turn everything into matrix notation

Intercept

$$h_\theta(x) = \theta^T x \qquad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \dots \\ \theta_n \end{pmatrix} \in \mathbb{R}^{n+1} \qquad x = \begin{pmatrix} x_0 \triangleq 1 \\ x_1 \\ \dots \\ x_n \end{pmatrix} \in \mathbb{R}^{n+1}$$

- Design matrix

又称

1st data point

$$X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m-1)} \\ x^{(m)} \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & \dots & \dots & \dots & x_n^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(m)} & \dots & \dots & \dots & x_n^{(m)} \end{bmatrix}$$

共有m个 data pt.

m^{th} data pt.

goal 目标是 $\min_{\theta} J$

??自己推一遍

Normal Equation

matrix calculus

令 $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $B = m \times n$

$(AB)^T = B^T A^T$

$\frac{\partial x^T B}{\partial x} = B$

见P285

$\frac{\partial x^T B}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}$ Jacobian

- Cost function

design matrix

背公式

$$J(\theta) = \frac{1}{2m}(X\theta - y)^T(X\theta - y)$$

- Throwing out the constant

$$J(\theta) = \theta^T X^T X\theta - 2(X\theta)^T y + y^T y$$

$\frac{\partial x^T b}{\partial x} = b$ $\frac{\partial x^T x}{\partial x} = 2x$

若$B=B^T$, $\frac{\partial x^T B x}{\partial x} = 2Bx$

- The "famous" normal equation

$$\frac{\partial J}{\partial \theta} = 2X^T X\theta - 2X^T y = 0$$

$$X^T X\theta = X^T y$$

regression coefficients

$$\theta = (X^T X)^{-1} X^T y$$

caveat $X^T X$ full rank

用 generalized inverse / pseudo inverse

monroe penrose

Residuals

inverse 来算

自学一下!!

$SS(fit) = \sum \iota^2$

- The residual should be normally distributed

R^2 表示

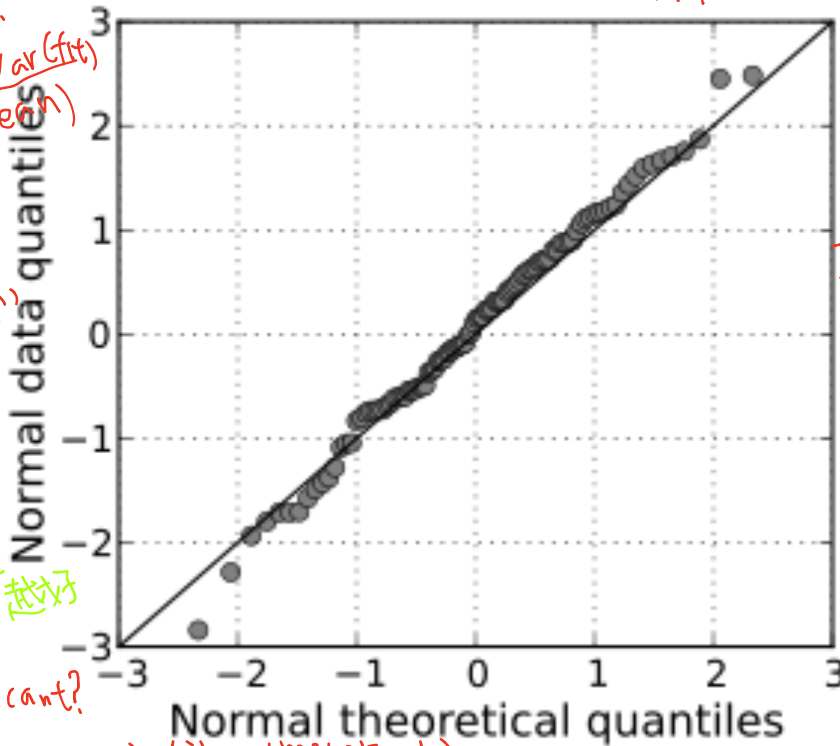
$$y^{(i)} = \theta^T x^{(i)} + \underbrace{\epsilon^{(i)}}_{\text{residual}}$$

$\rightarrow var(fit) = \frac{SS(fit)}{n}$

$R^2 = \frac{Var(mean) - Var(fit)}{Var(mean)}$

empirical quantiles

$\rightarrow Var(mean) = \frac{SS(mean)}{n}$



y

projection → y

without considering x

有 implicit assumption

$\epsilon \sim N$ 正态分布

利用 Q-Q plot 来判断是否正态

linreg 2个重点

R^2 越大 fit 越好

① R^2 r-squared $\in [0, 1]$

② Is R^2 significant?

F_test

$(SS(mean) - SS(fit)) / (\#parameter - 1)$

... sample points

Probabilistic Interpretation

F = (SS(me… − SS(fit)) / … n ~ sample po…

= varience explained by features / --- not ---

- Deriving the log-likilihood function

$$\epsilon^{(i)} = y^{(i)} - \theta^T x^{(i)} \sim \mathcal{N}\left(0, \sigma^2\right)$$

$$p\left(\epsilon^{(i)}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(\epsilon^{(i)}\right)^2}{2\sigma^2}\right)$$

normal density

$$\underbrace{p\left(y^{(i)} | x^{(i)}; \theta\right)}_{\text{likelihood}} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

Probabilistic Interpretation

- Deriving the log-likilihood function

$$L(\theta) = L(\theta; X, \vec{y}) = p(\vec{y}|X; \theta)$$

likelihood function

indep

$$L(\theta) = \prod_{i=1}^{m} p\left(y^{(i)}|x^{(i)}; \theta\right)$$

$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

max $L(\theta)$

求 θ

Probabilistic Interpretation

- Maximum Likelihood Estimator (MLE)

$$\ell(\theta) = \underline{\log L(\theta)}$$

log likelihood

$$= \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

$$= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

max

$$= m \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} \cdot \frac{1}{2} \underline{\sum_{i=1}^{m} \left(y^{(i)} - \theta^T x^{(i)}\right)^2}$$

min

least-square cost

Python Time!

- from sklearn import linear_model
- from statsmodels.api import OLS

ARMA(p,q)

auto regressive modle → stationary

time series model

- AR(p)

$$X_t = \mu + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

$|\varphi_i| < 1$

moving average model

- MA(q)

往前看q

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

- ARMA(p,q)

autoregressive moving average model with order (p,q)

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

order of

ARIMA(p,d,q)

trend

differencing.

integrated

- d is the degree of difference, e.g., d =1

build ARMA on Y_t

$$Y_t = X_t - X_{t-1}$$

$$Y_t = \mu + \varepsilon_t + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

- How about d = 2? 再令 $Z_t = Y_t - Y_{t-1}$

再 build ARMA on Z_t.

"detrend"

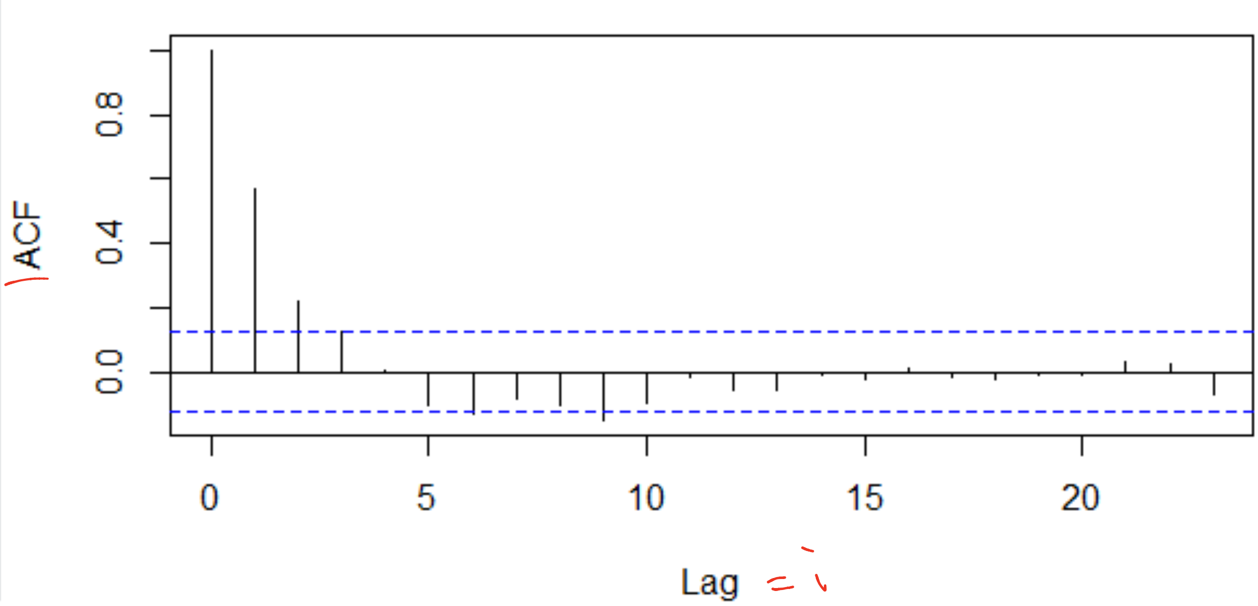
2 ways to

① linear reg.

② if difference stationary

ACF/PACF

auto correlation function $\Rightarrow$ determin p, q



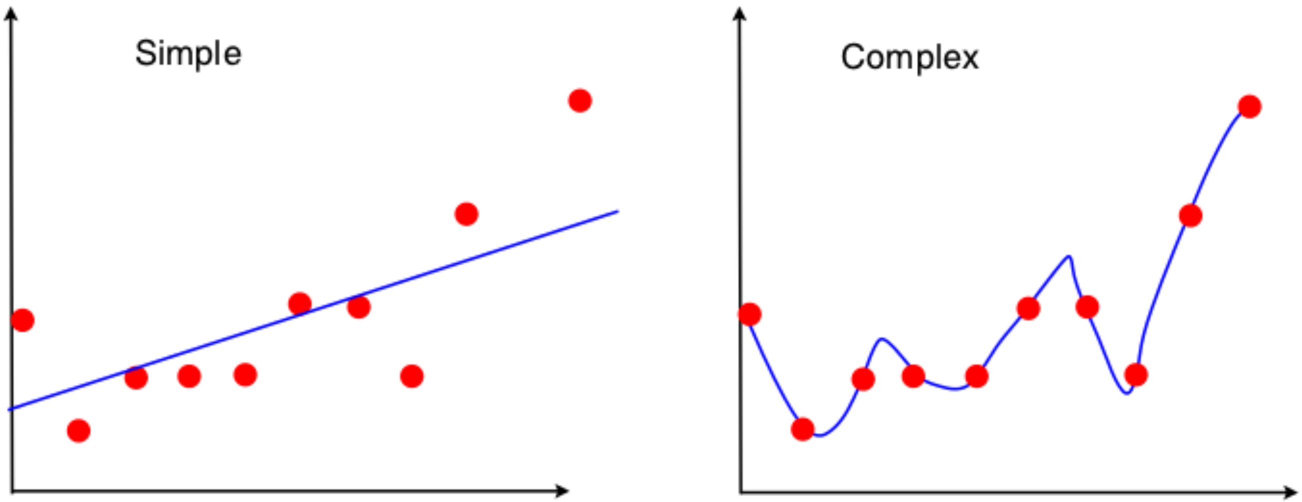
$= i$

$\text{corr}(X_t, X_{t-i})$

Python Time!

- statsmodels.tsa.arima_model

Bias-Variance Tradeoff



Kaggle Competition

All Competitions

Active | Completed | InClass

All Categories ▾ | Default Sort ▾

Jigsaw Multilingual Toxic Comment Classification — $50,000
Use TPUs to identify toxicity comments across multiple languages
Featured • a month to go • Code Competition • 862 Teams

M5 Forecasting - Accuracy — $50,000
Estimate the unit sales of Walmart retail goods
Featured • 2 months to go • 3589 Teams

M5 Forecasting - Uncertainty — $50,000
Estimate the uncertainty distribution of Walmart unit sales.
Featured • 2 months to go • 389 Teams

University of Liverpool - Ion Switching — $25,000
Identify the number of channels open at each time point
Research • 16 days to go • 2333 Teams

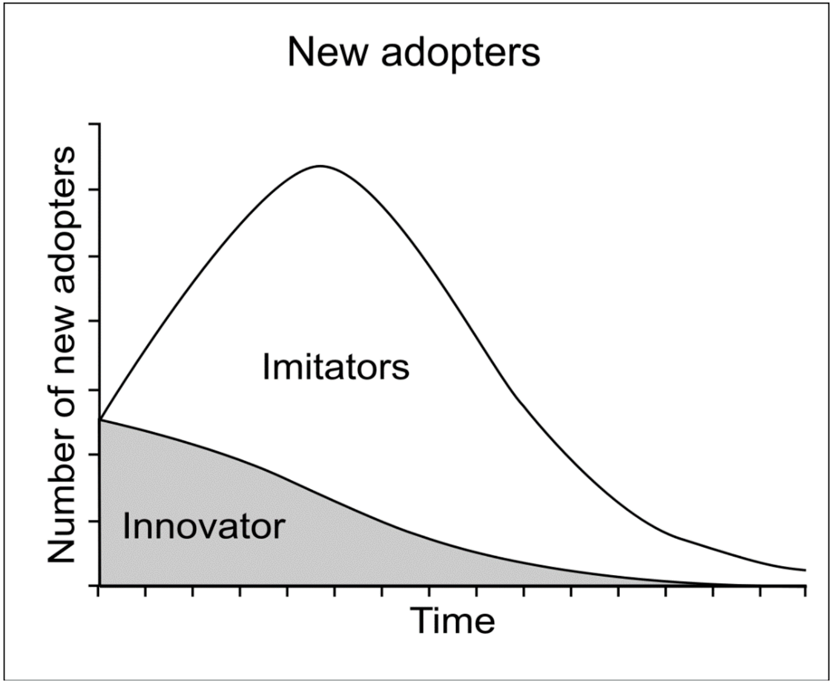
TReNDS Neuroimaging — $25,000
Multiscanner normative age and assessments prediction with brain function, structure, and connectivity
Research • 2 months to go • 275 Teams

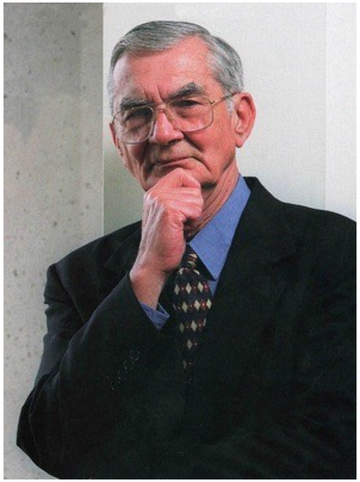
ALASKA2 Image Steganalysis — $25,000
Detect secret data hidden within digital images
Research • 2 months to go • 237 Teams

Prostate cANcer graDe Assessment (PANDA) Challenge — $25,000
Prostate cancer diagnosis using the Gleason grading system
Featured • 2 months to go • Code Competition • 309 Teams

Bass Diffusion Model

- How about projecting sales of new products?





Frank Bass

Bass Diffusion Model

Cumulative purchase probability of a random customer $F(t)$

Purchase probability at time t $f(t) = F'(t)$

The rate of purchase at time t (given no purchase so far)

$$\frac{f(t)}{1 - F(t)} = p + qF(t)$$

Coefficient of innovation

Coefficient of imitation

Bass Solution

$$\frac{dF/dt}{1-F} = p + qF$$

$$\frac{dF}{dt} = p + (q-p)F - qF^2$$

$$\int \frac{1}{p+(q-p)F-qF^2} dF = \int dt$$

$$\begin{aligned}
\frac{1}{(p+qF)(1-F)} &= \frac{A}{p+qF} + \frac{B}{1-F} \\
&= \frac{A - AF + pB + qFB}{(p+qF)(1-F)} \\
&= \frac{A + pB + F(qB - A)}{(p+qF)(1-F)}
\end{aligned}$$

$$A = q/(p+q)$$
$$B = 1/(p+q)$$

Bass Solution

$$\int \frac{1}{(p+qF)(1-F)} dF = \int dt$$

$$\int \left(\frac{A}{p+qF} + \frac{B}{1-F} \right) dF = t + c_1$$

$$\int \left(\frac{q/(p+q)}{p+qF} + \frac{1/(p+q)}{1-F} \right) dF = t + c_1$$

$$\frac{1}{p+q} \ln(p+qF) - \frac{1}{p+q} \ln(1-F) = t + c_1$$

$$\frac{\ln(p+qF) - \ln(1-F)}{p+q} = t + c_1$$

Boundary Condition

$$t = 0 \quad \Rightarrow \quad F(0) = 0$$

$$t = 0 \quad \Rightarrow \quad c_1 = \frac{\ln p}{p+q}$$

$$F(t) = \frac{p\left(e^{(p+q)t} - 1\right)}{pe^{(p+q)t} + q}$$

Bass Solution

Calibration

- Sales in any period are $s(t) = mf(t)$
- Cumulative sales up to time t are $S(t) = mF(t)$

$$\frac{s(t)/m}{1 - S(t)/m} = p + qS(t)/m$$

$$s(t) = [p + qS(t)/m][m - S(t)]$$

$$s(t) = \beta_0 + \beta_1 S(t) + \beta_2 S(t)^2 \quad (BASS)$$

$$\beta_0 = pm$$

$$\beta_1 = q - p$$

$$\beta_2 = -q/m$$

$$m = \frac{-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_0\beta_2}}{2\beta_1}$$

$$p = \frac{\beta_0}{m}; \quad q = -m\beta_2$$

Conduct a linear regression!

Python Time!

- from sklearn import linear_model
- from statsmodels.api import OLS

Discrete Choice Model

- Multinomial Logit Model

$$U_{ni} = V_{ni} + \epsilon_{ni}$$

Gumbel

$$f(x) = e^{-x} e^{-e^{-x}}$$

$$F(x) = e^{-e^{-x}}$$

$$\begin{aligned} P_{ni} &= \mathbb{P}\left(U_{ni} > U_{nj} \quad \forall j \neq i\right) \\ &= \mathbb{P}\left(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \quad \forall j \neq i\right) \end{aligned}$$

Discrete Choice Model

- Multinomial Logit Model

Gumbel

$$f(x) = e^{-x} e^{-e^{-x}}$$

$$F(x) = e^{-e^{-x}}$$

$$\begin{aligned} P_{ni} &= \int (P_{ni}|\epsilon_{ni}) \, f(\epsilon_{ni}) \, d\epsilon_{ni} \\ &= \int (P_{ni}|\epsilon_{ni}) \, e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}} \, d\epsilon_{ni} \\ &= \int \left(\prod_{j \neq i} e^{-e^{-(\epsilon_{ni} + V_{ni} - V_{nj})}} \right) e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}} \, d\epsilon_{ni} \end{aligned}$$

$$\Longrightarrow \qquad P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$