BAB III

TAKSIRAN KOEFISIEN KORELASI *POLYCHORIC* DUA TAHAP

Permasalahan dalam tugas akhir ini dibatasi hanya pada penaksiran besarnya koefisien korelasi *polychoric* dan tidak dilakukan pengujian terhadap koefisien korelasi *polychoric*. Taksiran koefisien korelasi *polychoric* dapat dicari dengan menggunakan metode taksiran dua tahap.

Misalkan *U* dan *V* adalah dua variabel *random* yang kontinu, selanjutnya misalkan *X* dan *Y* adalah variabel-variabel *ordinal* yang dibentuk dari variabel *random* kontinu *U* dan *V* maka hubungan antara variabel *ordinal* *X* dengan variabel *random* kontinu *U* yang membentuknya dapat ditulis sebagai berikut:

$$\begin{array}{lll} x = 1 & \text{jika} & u < a_1 \\ x = 2 & \text{jika} & a_1 \le u < a_2 \\ . & . & \\ . & . & \\ . & . & \\ x = I & \text{jika} & a_{I-1} \le u \end{array} \tag{3.1}$$

dimana, a_1 = batas atas kategori pertama variabel *X*, a_2 = batas atas kategori kedua variabel *X*,………, a_{I-1} = batas atas kategori ke- I-1 variabel

X, definisikan batas bawah kategori pertama variabel *X* dengan $a_0 = -\infty$ dan batas atas kategori ke- I variabel *X* dengan $a_I = \infty$. Demikian pula untuk hubungan antara variabel ordinal *Y* dengan variabel random kontinu *V* yang membentuknya, dengan b_1 = batas atas kategori pertama variabel *Y*, b_2 = batas atas kategori kedua variabel *Y*,………, b_{J-1} = batas atas kategori ke- J-1 variabel *Y*, serta $b_0 = -\infty$ dan $b_J = \infty$.

Hubungan antara variabel *ordinal X* dengan variabel *random* kontinu *U* yang membentuknya dapat diilustrasikan melalui gambar berikut :

Gambar 3.1 Hubungan Variabel *Ordinal X* dengan Variabel Kontinu *U*.

demikian juga hubungan antara variabel *ordinal Y* dengan variabel *random* kontinu *V* yang membentuknya dapat diilustrasikan melalui gambar berikut :

Gambar 3.2 Hubungan Variabel *Ordinal Y* dengan Variabel Kontinu *V*.

Ketika data yang teramati hanya berupa variabel-variabel ordinal *X* dan *Y* yang dibentuk oleh variabel random kontinu *U* dan *V*, menaksir kekuatan hubungan linier antara variabel random kontinu *U* dan *V* dengan menggunakan koefisien korelasi *pearson* tidak dimungkinkan. Salah satu cara untuk mengatasi permasalahan yang timbul dalam menghitung besarnya korelasi dua variabel random kontinu jika data yang teramati merupakan data ordinal yang dibentuk oleh kedua variabel random kontinu tersebut, adalah dengan menggunakan koefisien korelasi *polychoric* yang akan ditaksir dengan menggunakan metode taksiran dua tahap.

Metode taksiran dua tahap terdiri dari tahapan berikut :

1. Menaksir batas – batas untuk setiap kategori dari masing – masing data ordinal.
2. Menaksir koefisien korelasi *polychoric* dengan menggunakan taksiran batas – batas untuk setiap kategori dari masing – masing data *ordinal* yang diperoleh pada tahap pertama, melalui metode taksiran maksimum *likelihood*.

3.1 Tahap Pertama Metode Taksiran Dua Tahap

Seperti yang telah disebutkan sebelumnya, tahapan pertama dari metode taksiran dua tahap adalah menaksir batas – batas untuk setiap kategori dari masing – masing data *ordinal*. Pada tahapan ini, proporsi *marginal* sampel dalam setiap kategori digunakan untuk menaksir batas – batas untuk setiap kategori dari masing – masing data *ordinal*. Dengan demikian, distribusi dari variabel kontinu awal harus ditentukan atau diketahui.

Perhatikan tabel 2.1, misalkan variabel *ordinal* X berasal dari variabel kontinu U yang mempunyai fungsi distribusi $F_1(u)$ dan p.d.f $f_1(u)$ dan variabel *ordinal* Y berasal dari variabel kontinu V mempunyai fungsi distribusi $F_2(v)$ dan p.d.f $f_2(v)$ maka batas – batas kategori variabel X, $a_i, i = 1,......I - 1$ ditaksir dengan :

$$a_i = F_1^{-1}(P_1. + P_2. + ... + P_i.) \tag{3.1.1}$$

dan batas – batas kategori variabel Y, $b_j, j = 1,......J - 1$ ditaksir dengan :

$$b_j = F_2^{-1}(P._1 + P._2 + ... + P._j) \tag{3.1.2}$$

dimana: $P_i.$ = proporsi *marginal* kategori ke – i variabel *ordinal* X, dengan

$P_i. = \dfrac{n_i.}{n}$ dan $n_i. = \sum_{j=1}^{J} n_{ij}$.

$P._j$ = proporsi *marginal* kategori ke -j variabel *ordinal* Y, dengan

$$P_{\cdot j} = \frac{n_{\cdot j}}{n} \text{ dan } n_{\bullet j} = \sum_{i=1}^{I} n_{ij} .$$

F_1^{-1} = invers dari fungsi distribusi $F_1(u)$.

F_2^{-1} = invers dari fungsi distribusi $F_2(v)$.

Dalam metode taksiran dua tahap yang dibahas dalam tugas akhir ini, diasumsikan distribusi gabungan dari variabel *U* dan *V* adalah normal bivariat standar maka dapat ditunjukkan *U* dan *V* masing – masing berdistribusi $N(0,1)$ (lihat lampiran 3). Walaupun demikian, akan ditunjukkan (melalui simulasi pada bab 4) bahwa taksiran koefisien korelasi *polychoric* yang didapat *robust* terhadap asumsi tersebut. Karena *U* dan *V* masing – masing berdistribusi $N(0,1)$ maka batas – batas kategori variabel *X*, $a_i, i = 1,......I-1$ ditaksir dengan :

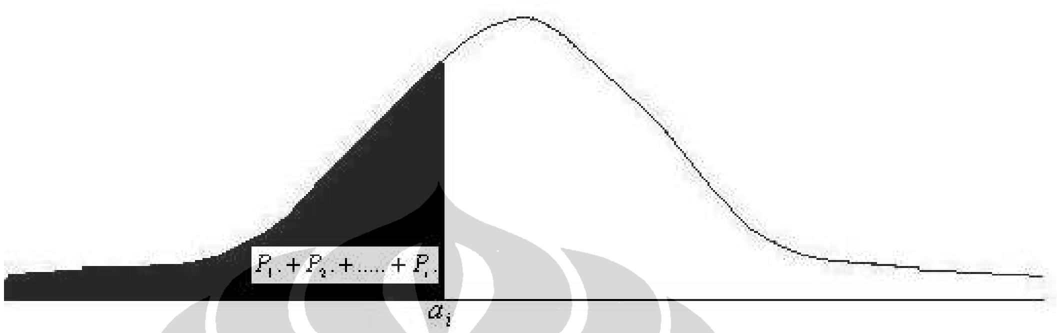
$$a_i = \Phi_1^{-1}(P_1. + P_2. + ... + P_i.) \qquad (3.1.3)$$

dan batas – batas kategori variabel *Y*, $b_j, j = 1,......J-1$ ditaksir dengan :

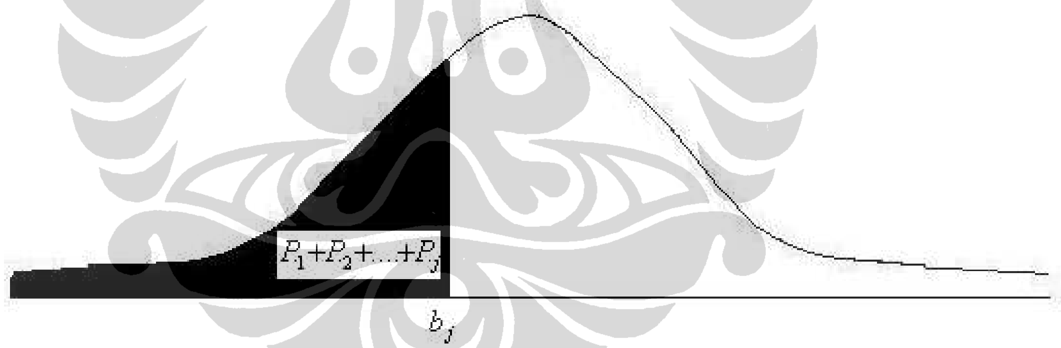
$$b_j = \Phi_1^{-1}\left(P._1 + P._2 + ... + P._j\right) \qquad (3.1.4)$$

dimana : Φ_1^{-1} = invers fungsi distribusi normal univariat standar.

Proses penaksiran batas – batas kategori untuk masing – masing variabel *ordinal X* dan *Y* dapat diilustrasikan melalui gambar berikut :



Gambar 3.1.1.1 Penaksiran Batas Atas Kategori ke – i Variabel *Ordinal X*.



Gambar 3.1.1.2 Penaksiran Batas Atas Kategori ke – j Variabel *Ordinal Y*.

3.2 Tahap Kedua Metode Taksiran Dua Tahap

Penaksiran besarnya koefisien korelasi *polychoric* pada tahap kedua pada dasarnya menggunakan metode taksiran maksimum *likelihood*, yang membutuhkan distribusi gabungan dari variabel *U* dan *V*. Misalkan (X_1, Y_1), (X_2, Y_2), …, (X_n, Y_n) merupakan sampel random bivariat dari variabel *ordinal* *X* dan *Y* maka fungsi *likelihood* dari sampel *random* bivariat ini adalah :

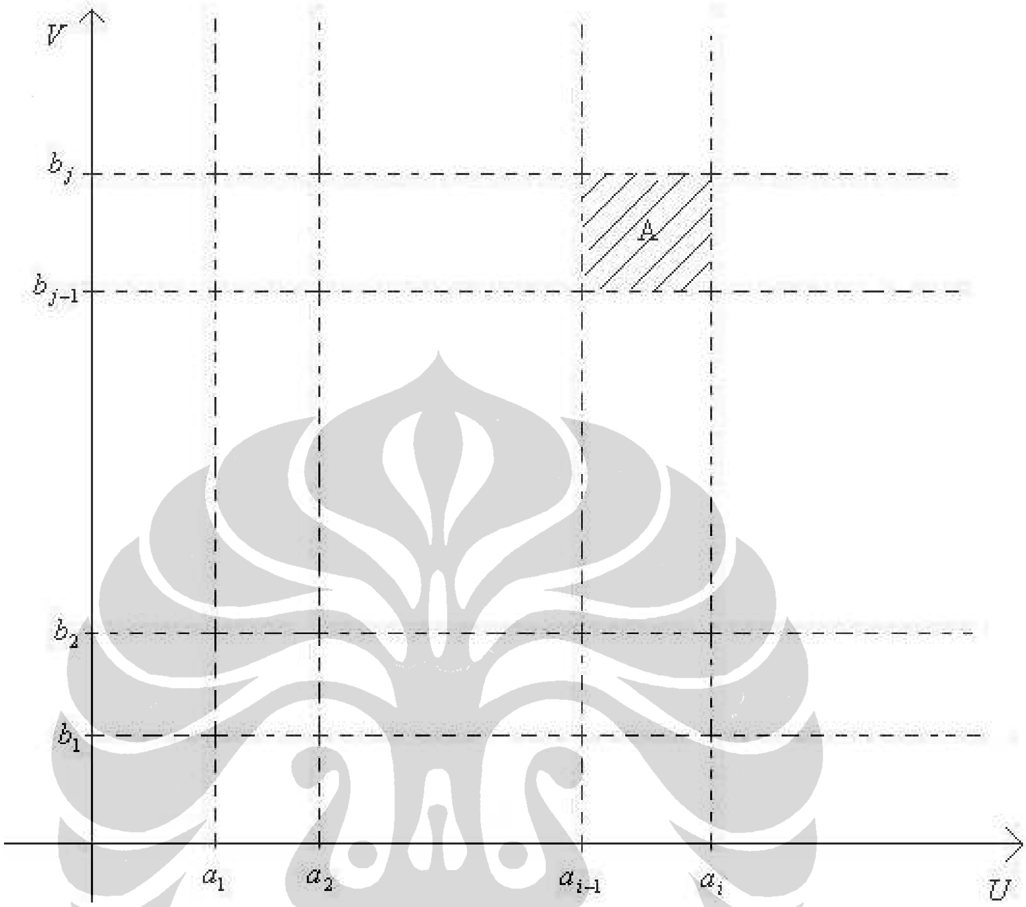
$$\begin{aligned} L &= \pi_{11}^{n_{11}} \pi_{12}^{n_{12}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \pi_{IJ}^{n_{ij}} \\ &= \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{\text{ij}}^{n_{ij}} \end{aligned} \qquad (3.2.1)$$

dimana

π_{ij} = probabilitas suatu observasi jatuh pada sel (*i*, *j*), *i* = 1,……,I, *j* = 1,…..,J dari tabel kontingensi variabel *ordinal* *X* dan *Y*.

n_{ij} = banyaknya observasi yang jatuh pada sel (*i*, *j*) *i* = 1,……,I, *j* = 1,…..,J dari tabel kontingensi variabel *ordinal* *X* dan *Y*.

Untuk mencari besarnya probabilitas suatu observasi jatuh pada sel (*i*, *j*), *i* = 1,……,I, *j* = 1,….., J dari tabel kontingensi variabel *ordinal* *X* dan *Y* , perhatikan gambar berikut :



Gambar 3.2.1 Ilustrasi Perhitungan π_{ij} .

Probabilitas suatu observasi jatuh pada sel (*i*, *j*) akan sama dengan :

$$\pi_{ij} = \int_{b_{j-1}}^{b_j} \int_{a_{i-1}}^{a_i} f(u,v)\, du\, dv \qquad (3.2.2)$$

dimana : $f(u,v)$ adalah p.d.f gabungan dari variabel *U* dan *V*.

Jika $F(u,v)$ adalah fungsi distribusi gabungan dari *U* dan *V* maka :

$$\pi_{ij} = F(a_i, b_j) - F(a_{i-1}, b_j) - F(a_i, b_{j-1}) + F(a_{i-1}, b_{j-1}) \qquad (3.2.3)$$

Karena distribusi gabungan *U* dan *V* diasumsikan normal bivariat standar maka :

$$\pi_{ij} = \int_{b_{j-1}}^{b_j} \int_{a_{i-1}}^{a_i} \Phi_2(u,v)\, du\, dv$$

$$\pi_{ij} = \int_{b_{j-1}}^{b_j} \int_{a_{i-1}}^{a_i} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ \frac{-1}{2(1-\rho^2)} \left(u^2 - 2\rho uv + v^2 \right) \right\} du\, dv \qquad (3.2.4)$$

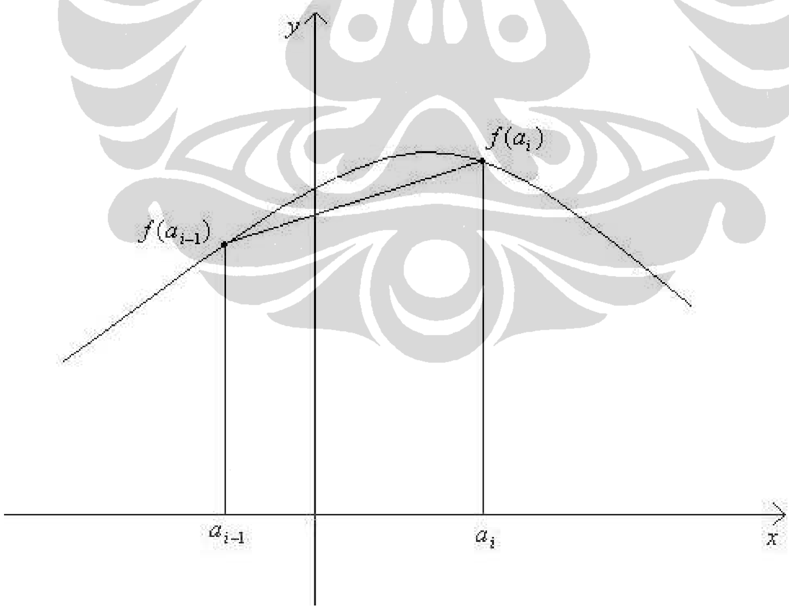
dengan $\Phi_2(u,v)$ adalah p.d.f normal bivariat standar (lihat lampiran 3).

Nilai π_{ij} pada persamaan (3.2.4) biasanya sulit atau tidak bisa didapat secara analitis. Oleh sebab itu, nilai π_{ij} pada persamaan (3.2.4) akan dihitung melalui pendekatan numerik, salah satu metode yang dapat digunakan adalah metode *trapezoid rule*.

Trapezoid rule adalah salah satu metode pendekatan numerik untuk menghitung integral berhingga $\int_{a_{i-1}}^{a_i} f(x)\,dx$, dengan cara mengaproksimasi daerah di bawah fungsi $f(x)$ dengan sebuah trapesium, kemudian menghitung luas trapesium ini sebagai aproksimasi dari $\int_{a_{i-1}}^{a_i} f(x)\,dx$, yaitu melalui formula berikut :

$$\int_{a_{i-1}}^{a_i} f(x)\,dx \approx \frac{a_i - a_{i-1}}{2}\left(f(a_{i-1}) + f(a_i)\right) \qquad (3.2.5)$$

Metode *trapezoid rule* dapat diilustrasikan melalui gambar berikut :



Gambar 3.2.2 Ilustrasi Metode *Trapezoid Rule* .

Karena terdapat beberapa π_{ij} yang dihitung melalui persamaan (3.2.4) merupakan integral tak berhingga maka sebelum menggunakan metode *trapezoid rule*, batas – batas integral yang tak berhingga tersebut harus ditransformasi terlebih dahulu agar menjadi berhingga. Misalkan akan dihitung π_{11}, berdasarkan persamaan (3.2.4) :

$$\pi_{11} = \int_{-\infty}^{b_1}\int_{-\infty}^{a_1} \Phi_2(u,v)\, du\, dv$$

$$\pi_{11} = \int_{-\infty}^{b_1}\int_{-\infty}^{a_1} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left(u^2 - 2\rho uv + v^2\right)\right\} \; du\, dv$$

Dengan menganggap *v* sebagai konstanta, $\Phi_2(u,v)$ akan diintegralkan terlebih dahulu terhadap *u*. Pandang transformasi $u = a_1 - \frac{1-s}{s}$ jika $u == -\infty$ maka $s = 0$ dan jika $u = a_1$ maka $s = 1$, sehingga diperoleh

$\int_0^1 \Phi_2(\mathrm{s},\mathrm{v}) \frac{1}{\mathrm{s}^2}\, ds$, sebut $\Phi_2(s,v)\frac{1}{s^2} = g_1(s,v)$. Dengan metode *trapezoid rule*, dapat dihitung $\int_0^1 g_1(s,v)\, ds$ yaitu :

$$\int_0^1 g_1(s,v)\, ds \approx \frac{1}{2}\left(g_1(0,v) + g_1(1,v)\right)$$

hasil dari $\int_0^1 g_1(s,v)\, ds$ merupakan fungsi dari *v*, sebut $g_2(v)$. Selanjutnya $g_2(v)$ akan diintegralkan terhadap *v* untuk memperoleh π_{11}, sebagai berikut :

$\pi_{11} = \int_{-\infty}^{b_1} g_2(v)\, dv$, misal $v = b_1 - \frac{1-t}{t}$ jika $v = -\infty$ maka $t = 0$ dan jika $v = b_1$ maka $t = 1$, sehingga diperoleh :

$\pi_{11} = \int_{-\infty}^{b_1} g_2(v)\, dv = \int_0^1 g_2\left(b_1 - \frac{1-t}{t}\right)\frac{1}{t^2}\, dt$, sebut $g_2\left(b_1 - \frac{1-t}{t}\right)\frac{1}{t^2} = g_3(t)$. Dengan metode *trapezoid rule*, dapat dihitung $\int_0^1 g_3(t)\, dt$ yaitu :

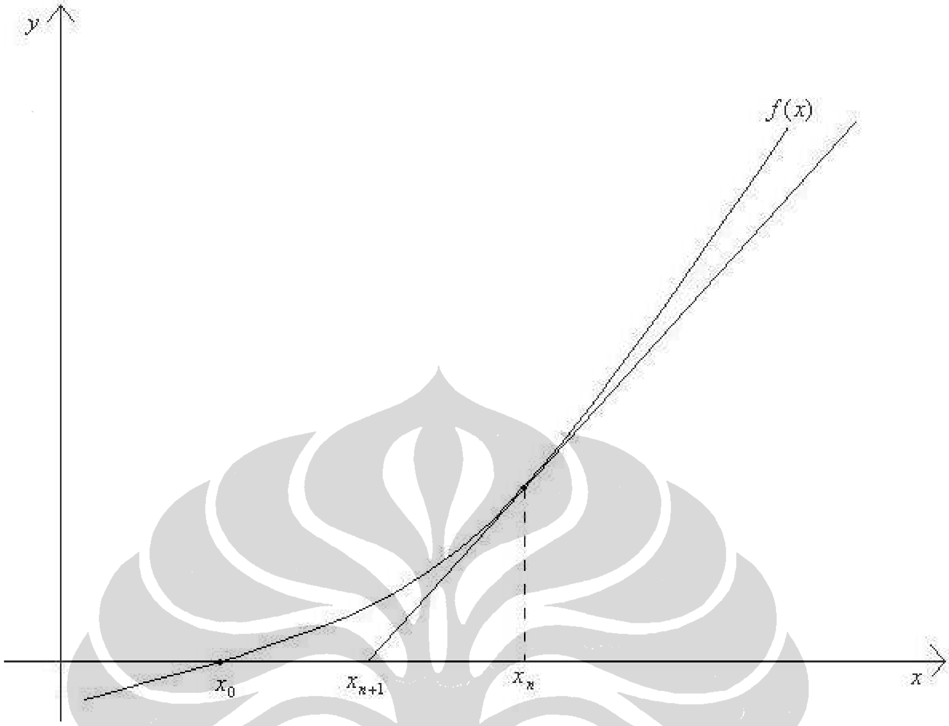
$\int_0^1 g_3(t)\, dt \approx \frac{1}{2}\left(g_3(0) + g_3(1)\right)$. Dengan demikian, diperoleh besarnya $\pi_{11} = \int_0^1 g_3(t)\, dt$ yang merupakan fungsi dari ρ (parameter distribusi normal bivariat standar). Dengan cara yang sama dapat dihitung $\pi_{12},...,\pi_{1J},\pi_{21},....,\pi_{I1},\pi_{2J},....,\pi_{I-1J},\pi_{I2},....,\pi_{IJ}$, untuk perhitungan yang memuat integral seperti berikut : $\int_a^\infty f(x)\, dx$, batas integralnya dapat ditransformasi dengan $x = a + \frac{1-t}{t}$.

Dengan menyubstitusikan besar π_{ij} pada persamaan (3.2.4) ke persamaan (3.2.1) maka fungsi *likelihood* (3.2.1) merupakan fungsi dari parameter ρ. Fungsi *loglikelihood* dari sampel *random* bivariat (X_1, Y_1), (X_2, Y_2), ………, (X_n, Y_n) dapat ditulis sebagai berikut :

$$\ln L(\rho) = l(\rho) = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \ln \pi_{ij} \tag{3.2.6}$$

Taksiran maksimum *likelihood* dari ρ diperoleh dengan mencari nilai ρ yang memaksimumkan fungsi *loglikelihood* (3.2.6), yaitu fungsi *loglikelihood* (3.2.6) diturunkan terhadap ρ dan disamakan dengan nol. Perhitungan taksiran maksimum *likelihood* dari ρ ini biasanya sulit atau tidak dapat diselesaikan secara analitis tetapi dapat diselesaikan melalui pendekatan numerik, salah satu metode yang dapat digunakan adalah metode *newton raphson*.

Metode *newton raphson* adalah salah satu metode numerik untuk mencari aproksimasi akar atau pembuat nol dari fungsi bernilai real. Misalkan $f(x)$ adalah fungsi bernilai real, cara kerja dari metode *newton raphson* adalah dengan menebak akar dari fungsi $f(x)$ untuk dijadikan sebagai nilai awal (misal x_n). Kemudian buat garis singgung dari $f(x)$ di $(x_n, f(x_n))$ dan hitung juga titik potong dengan sumbu x dari garis singgung tersebut . Misalkan x_{n+1} adalah titik potong dengan sumbu x dari garis singgung dari $f(x)$ di $(x_n, f(x_n))$, x_{n+1} ini merupakan aproksimasi akar $f(x)$ yang lebih baik dari x_n. Untuk lebih jelasnya, perhatikan gambar berikut :



Gambar 3.2.3 Ilustrasi Metode *Newton Raphson*.

Karena turunan $f(x)$ di titik x_n merupakan kemiringan garis singgung dari $f(x)$ di titik x_n maka

$$\begin{aligned} & f'(x_n) = \frac{\Delta y}{\Delta x} = \frac{f(x_n) - 0}{x_n - x_{n+1}} = \frac{0 - f(x_n)}{(x_{n+1} - x_n)} \\ \Leftrightarrow\ & f'(x_n)(x_{n+1} - x_n) = -f(x_n) \\ \Leftrightarrow\ & x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (3.2.7) \end{aligned}$$

proses ini dilakukan secara iterasi sampai x_n konvergen ke nilai tertentu.

Kekonvergenan dari metode *newton raphson* dapat dijelaskan melalui teorema berikut :

Teorema 3.1 :

Misalkan $f \in C^2[a,b]$. Jika $p \in [a,b]$ sedemikian sehingga $f(p) = 0$ dan $f\,'(p) \neq 0$ maka ada $\delta > 0$ sedemikian sehingga metode newton raphson menghasilkan barisan $\{p_n\}_{n=1}^{\infty}$ yang konvergen ke p untuk sembarang nilai awal $p_0 \in [p-\delta, p+\delta]$. (bukti di lampiran 4).

Metode *newton raphson* dapat digunakan untuk mencari nilai maksimum dari fungsi *loglikelihood* (3.2.6) dengan memandang $l\,'(\rho)$ sebagai $f(x)$ sehingga akan diperoleh taksiran maksimum *likelihood* dari ρ dengan menggunakan persamaan (3.2.7) :

$$\rho_{n+1} = \rho_n - \frac{l\,'(\rho_n)}{l\,''(\rho_n)} \tag{3.2.8}$$

Taksiran maksimum *likelihood* dari ρ inilah yang disebut dengan koefisien korelasi *polychoric* antara dua variabel *ordinal X* dan *Y* yang teramati. Karena fungsi $l\,'(\rho)$ memiliki domain ruang parameter distribusi normal bivariat standar ($\rho \in [-1,1]$) sedemikian sehingga $l\,'(\rho) = 0$ dan $l\,''(\rho) \neq 0$ maka berdasarkan teorema 3.1 ρ_{n+1} akan konvergen ke ρ. Dengan kata lain, taksiran koefisien korelasi *polychoric* pada persamaan (3.2.8) akan konvergen ke nilai koefisien korelasi populasi, yaitu koefisien korelasi *polychoric* akan bernilai antara –1 sampai dengan 1.