Xyba and rilo_chand Project

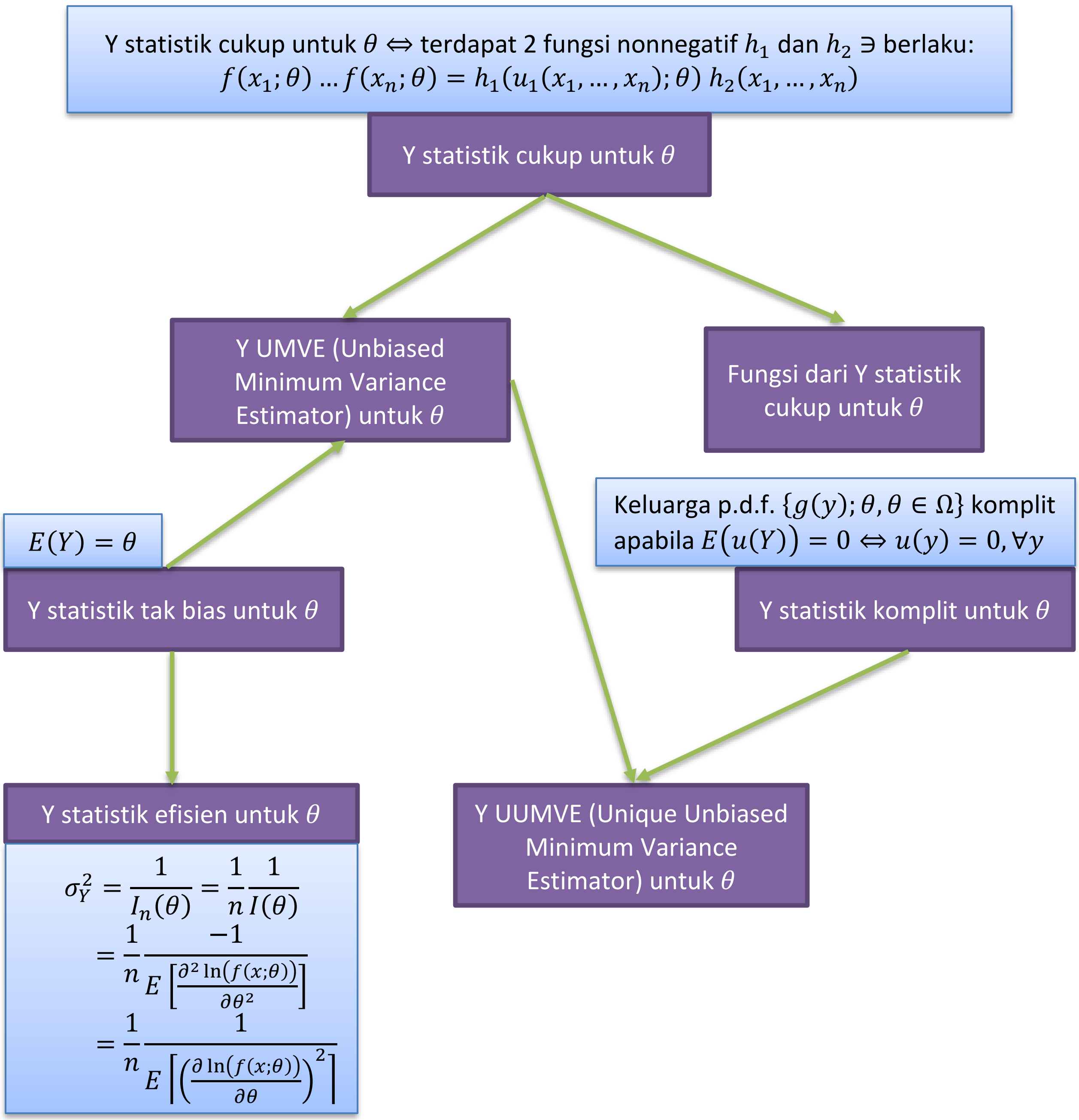
Statistika Matematika 2

Rangkuman UAS

1. This document is created by Xyba and rilo_chand, students of Mathematics University of Indonesia Batch 2016
2. Should there be any mistakes or feedbacks you'd like to give, please contact either one of us
3. Last Updated: 30/05/2018

Statistik Penaksir Satu Parameter

Misal $X_1, \dots, X_n$ adalah sampel random berukuran n dari distribusi yang mempunyai p.d.f. $f(x;\theta), \theta \in \Omega$. Misal $Y = u(X_1, \dots, X_n)$ adalah statistik untuk θ yang mempunyai p.d.f. $g(y;\theta)$.



6.2. : Interval Kepercayaan untuk Mean

1. Distribusi Normal

   $X_1, \ldots, X_n \sim NIID(\mu, \sigma^2)$

   a. Jika σ diketahui namun μ tidak diketahui

   $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{X-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Leftrightarrow Z \sim N(0,1)$

   Interval random untuk μ diberikan oleh:

   $$\left(\bar{X} - \frac{z\sigma}{\sqrt{n}}, \bar{X} + \frac{z\sigma}{\sqrt{n}}\right)$$

   Dan probabilitasnya adalah:

   $$\Pr\left(\bar{X} - \frac{z\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{z\sigma}{\sqrt{n}}\right) = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1$$

   Setelah melakukan percobaan random, maka interval kepercayaan $(1-\alpha)100\%$ untuk μ adalah:

   $$\left(\bar{x} - \frac{z\sigma}{\sqrt{n}}, \bar{x} + \frac{z\sigma}{\sqrt{n}}\right)$$

   dimana $z = z_{\alpha/2}$.

   b. Jika σ tidak diketahui dan μ tidak diketahui

   Untuk n kecil, gunakan distribusi t

   $$T = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{nS^2/\sigma^2(n-1)}} = \frac{\bar{X}-\mu}{S/\sqrt{n-1}} \sim t_{n-1}$$

   Interval random untuk μ diberikan oleh:

   $$\left(\bar{X} - \frac{bS}{\sqrt{n-1}}, \bar{X} + \frac{bS}{\sqrt{n-1}}\right)$$

   Dan probabilitasnya adalah:

   $$\Pr\left(\bar{X} - \frac{bS}{\sqrt{n-1}} < \mu < \bar{X} + \frac{bS}{\sqrt{n-1}}\right) = \Pr(T \le t) - \Pr(T \le -t) = 2\Pr(T \le t) - 1$$

   Setelah melakukan percobaan random, maka interval kepercayaan untuk μ adalah:

   $$\left(\bar{x} - \frac{bs}{\sqrt{n-1}}, \bar{x} + \frac{bs}{\sqrt{n-1}}\right)$$

   dimana $b = t_{\alpha/2}$.

   Untuk n besar, gunakan Central Limit Theorem atau Standardisasi, lalu gunakan rumus pada poin (a) jika σ diketahui dan poin (b) jika σ tidak diketahui.

2. Distribusi selain Normal

   a. Jika σ diketahui namun μ tidak diketahui

   Gunakan fakta bahwa:

   $$S^2 \xrightarrow{P} \sigma^2, \qquad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1), \qquad \sqrt{\frac{nS^2}{(n-1)\sigma^2}} \xrightarrow{P} 1$$

Maka:

$$\frac{\bar{X}-\mu}{S/\sqrt{n-1}} \xrightarrow{D} N(0,1)$$

Dalam hal ini ukuran sampel harus cukup besar apabila distribusi tidak simetris.

3. Distribusi Binomial

Misalkan $Y \sim b(n,p)$ dengan n diketahui cukup besar. Akan dicari interval kepercayaan untuk p.

Pada subbab 5.5, telah dibuktikan bahwa variabel random

$\dfrac{Y-np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1)$ dan $n\dfrac{Y}{n}\left(1-\dfrac{Y}{n}\right) \xrightarrow{P} np(1-p)$, sehingga

$$Z = \frac{\frac{Y-np}{\sqrt{np(1-p)}}}{\sqrt{\frac{n\frac{Y}{n}\left(1-\frac{Y}{n}\right)}{np(1-p)}}} = \frac{Y-np}{\sqrt{n\frac{Y}{n}\left(1-\frac{Y}{n}\right)}} \xrightarrow{D} N(0,1)$$

Untuk memperoleh interval random, substitusi Z ke

$$\Pr\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha$$

Interval random untuk p diberikan oleh

$$\left(\frac{Y}{n} - z_{\frac{\alpha}{2}}\sqrt{n\frac{Y}{n}\left(1-\frac{Y}{n}\right)}, \frac{Y}{n} + z_{\frac{\alpha}{2}}\sqrt{n\frac{Y}{n}\left(1-\frac{Y}{n}\right)}\right)$$

Setelah melakukan percobaan random, diperoleh interval kepercayaan $(1-\alpha)100\%$ untuk p adalah

$$\left(\frac{y}{n} - z_{\frac{\alpha}{2}}\sqrt{n\frac{y}{n}\left(1-\frac{y}{n}\right)}, \frac{y}{n} + z_{\frac{\alpha}{2}}\sqrt{n\frac{y}{n}\left(1-\frac{y}{n}\right)}\right)$$

6.3. : Interval Kepercayaan untuk Selisih Mean

1. Distribusi Normal

$X_1, \dots, X_n \sim NIID(\mu_1, \sigma_1^2)$

$Y_1, \dots, Y_m \sim NIID(\mu_2, \sigma_2^2)$

Kedua sampel saling bebas satu sama lain.

a. σ_1 dan σ_2 diketahui.

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right), \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right) \Rightarrow \bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right)$$

Maka

$$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1)$$

Substitusi Z ke

$$\Pr\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Sehingga diperoleh interval random untuk $\mu_1 - \mu_2$ adalah

$$\left((\bar{X} - \bar{Y}) - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, (\bar{X} - \bar{Y}) + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

Setelah melakukan percobaan random, diperoleh interval random $(1-\alpha)100\%$ untuk $\mu_1 - \mu_2$ adalah

$$\left((\bar{x} - \bar{y}) - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, (\bar{x} - \bar{y}) + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

b. $\sigma_1^2 = \sigma_2^2$ tetapi keduanya tidak diketahui.
Gunakan variabel random

$$T = \frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}}{\sqrt{\frac{nS_1^2}{\sigma_1^2} + \frac{mS_2^2}{\sigma_2^2} / (n+m-2)}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{nS_1^2+mS_2^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t_{(n+m-2)}$$

Substitusi T ke
$\Pr(-b < T < b) = 1 - \alpha$
untuk memperoleh interval random untuk $\mu_1 - \mu_2$, yaitu

$$\left((\bar{X} - \bar{Y}) - b \sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)}, (\bar{X} - \bar{Y}) + b \sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)} \right)$$

Setelah melakukan percobaan random, diperoleh interval kepercayaan $(1-\alpha)100\%$ untuk $\mu_1 - \mu_2$ yaitu

$$\left((\bar{x} - \bar{y}) - b \sqrt{\frac{ns_1^2 + ms_2^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)}, (\bar{x} - \bar{y}) + b \sqrt{\frac{ns_1^2 + ms_2^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)} \right)$$

dimana $b = t_{\alpha/2}$.

2. Distribusi Binomial
Misal $Y_1 \sim b(n_1, p_1)$ dan $Y_2 \sim b(n_2, p_2)$ saling bebas. n_1 dan n_2 diketahui dan cukup besar. Akan dicari interval kepercayaan untuk $p_1 - p_2$.
Perhatikan bahwa mean dan variansi dari $\frac{Y_1}{n_1} - \frac{Y_2}{n_2}$ berturut-turut adalah $p_1 - p_2$ dan $\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$. Maka, variabel random

$$\frac{\frac{Y_1}{n_1} - \frac{Y_2}{n_2} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \xrightarrow{D} N(0,1)$$

Kemudian, dapat dibuktikan bahwa jika $\frac{n_1}{n_2} = c > 0$ maka variabel random

$$\frac{\frac{\frac{Y_1}{n_1}\left(1-\frac{Y_1}{n_1}\right)}{n_1}+\frac{\frac{Y_2}{n_2}\left(1-\frac{Y_2}{n_2}\right)}{n_2}}{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}} \xrightarrow{P} 1$$

Sehingga variabel random

$$Z = \frac{\frac{\frac{Y_1}{n_1}-\frac{Y_2}{n_2}-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}}{\sqrt{\frac{\frac{\frac{Y_1}{n_1}\left(1-\frac{Y_1}{n_1}\right)}{n_1}+\frac{\frac{Y_2}{n_2}\left(1-\frac{Y_2}{n_2}\right)}{n_2}}{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}} = \frac{\frac{Y_1}{n_1}-\frac{Y_2}{n_2}-(p_1-p_2)}{\sqrt{\frac{\frac{Y_1}{n_1}\left(1-\frac{Y_1}{n_1}\right)}{n_1}+\frac{\frac{Y_2}{n_2}\left(1-\frac{Y_2}{n_2}\right)}{n_2}}} \xrightarrow{D} N(0,1)$$

Substitusi Z ke

$$\Pr\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Sehingga diperoleh interval random untuk $p_1 - p_2$ adalah

$$\left(\frac{Y_1}{n_1}-\frac{Y_2}{n_2}-z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{Y_1}{n_1}\left(1-\frac{Y_1}{n_1}\right)}{n_1}+\frac{\frac{Y_2}{n_2}\left(1-\frac{Y_2}{n_2}\right)}{n_2}}, \frac{Y_1}{n_1}-\frac{Y_2}{n_2}+z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{Y_1}{n_1}\left(1-\frac{Y_1}{n_1}\right)}{n_1}+\frac{\frac{Y_2}{n_2}\left(1-\frac{Y_2}{n_2}\right)}{n_2}}\right)$$

Setelah melakukan percobaan random, diperoleh interval random $(1-\alpha)100\%$ untuk $p_1 - p_2$ adalah

$$\left(\frac{y_1}{n_1}-\frac{y_2}{n_2}-z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{y_1}{n_1}\left(1-\frac{y_1}{n_1}\right)}{n_1}+\frac{\frac{y_2}{n_2}\left(1-\frac{y_2}{n_2}\right)}{n_2}}, \frac{y_1}{n_1}-\frac{y_2}{n_2}+z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{y_1}{n_1}\left(1-\frac{y_1}{n_1}\right)}{n_1}+\frac{\frac{y_2}{n_2}\left(1-\frac{y_2}{n_2}\right)}{n_2}}\right)$$

6.4. : Tes dari Hipotesis Statistika

Definisi 1

Hipotesis statistika adalah suatu penegasan tentang distribusi dari satu atau lebih variabel random.

Jika hipotesis statistika secara lengkap menspesifikasikan suatu distribusi, maka hipotesis statistika ini disebut hipotesis statistika sederhana.

Jika tidak secara lengkap menspesifikasikan suatu distribusi maka hipotesis statistika ini disebut hipotesis statistika majemuk/komposit.

Contoh :

<u>Sederhana</u>

$H_0: \theta = 75$

$H_1: \theta = 77$

<u>Komposit</u>

$H_0: \theta \leq 75$

$H_1: \theta > 75$

Dalam menentukan hipotesis statistika, apa yang menjadi dugaan kita dijadikan sebagai H_1 (sering disebut hipotesis alternatif).
Jika hipotesis statistikanya adalah komposit maka di dalam H_1 tidak boleh ada tanda =, dan tanda = diletakkan di H_0.

Definisi 2
Tes dari hipotesis statistik adalah suatu aturan yang mana jika hasil dari suatu percobaan telah diperoleh, maka hasil tersebut dapat membantu dalam memutuskan apakah menerima atau menolak suatu hipotesis yang telah ditentukan di awal.

Dalam membentuk suatu uji hipotesis, hal yang kita lakukan adalah menentukan kriteria penolakan H_0, dengan cara mempartisi $\mathcal{A}$ menjadi dua himpunan C dan C^* yang saling disjoint.
Misal $X_1, \dots, X_n$ sampel random yang memiliki nilai $x_1, \dots, x_n$.
Jika $(x_1, \dots, x_n) \in C$, maka kita menolak H_0. C kita sebut sebagai daerah kritis.

Kekuatan uji kita ditentukan oleh *power function,* yaitu probabilitas penolakan H_0, disimbolkan sebagai $K(\theta)$, atau

$$K(\theta) = P(C)$$

Kita menginginkan *power function* ketika H_0 benar adalah kecil, sementara *power function* ketika H_0 salah adalah besar.
Nilai maksimum (atau tepatnya adalah supremum) dari *power function* dari tes jika H_0 benar disebut sebagai tingkat signifikansi dari tes (atau ukuran dari daerah kritis C). Disimbolkan sebagai α.

Kesalahan ketika menolak H_0 padahal H_0 benar disebut kesalahan tipe I. Sedangkan kesalahan ketika menerima H_0 padahal H_0 salah disebut kesalahan tipe II.
Probabilitas kesalahan tipe II = 1 – power dari tes jika H_0 benar.

7.6. : Fungsi dari sebuah Parameter
Misal kita tidak tertarik dengan θ namun fungsi dari θ, kita ingin mencari UMVE untuk fungsi dari θ.
Misal X adalah sebuah variabel random dengan p.d.f. $f(x;\theta), \theta \in \Omega$.
Misal kita ingin mengestimasi bagian dari probabilitas distribusi ini pada, atau ke kiri, dari suatu titik c, dengan kata lain, $\Pr(X \leq c)$. Artinya kita mencari UMVE dari $F(c;\theta)$ dimana $F(x;\theta)$ adalah fungsi distribusi dari X.
Tulis $\Pr(X \leq c) = \Phi(c - \theta)$, dimana c adalah konstanta yang tetap.
Apabila $X_1, \dots, X_n \sim NIID(\mu, \sigma^2)$ dimana σ^2 diketahui, maka $E[u(X_1)|\bar{X} = \bar{x}] = \varphi(\bar{x})$ adalah UMVE untuk $\Phi(c - \theta)$. Lebih lanjut,

$$\varphi(\bar{X}) = \Phi\left[\sqrt{n}(c - \bar{X})/\sqrt{n-1}\right]$$

adalah UUMVE untuk $\Phi(c - \theta)$.

e.g.: (Soal 7.41 dari buku edisi 5)
Misal $X_1, X_2, \dots, X_n$ menotasikan sampel random dari distribusi $N(0, \theta)$. Maka $Y = \sum X_i^2$ adalah statistik cukup dan komplit untuk θ. Tentukan UMVE dari θ^2.

Jawab:
Diketahui $X_1, \dots, X_n \sim NIID(0, \theta)$.
Diberikan $Y = \sum_{i=1}^{n} X_i^2$ adalah statistik cukup dan komplit untuk θ.
Berdasarkan teorema , $\exists$ statistik yang merupakan UUMVE untuk θ yang merupakan fungsi dari Y yang dapat kita cari sebagai berikut.
Karena $X \sim N(0, \theta) \Rightarrow \frac{X}{\sqrt{\theta}} \sim N(0,1) \Rightarrow \frac{X^2}{\theta} \sim \chi^2_{(1)} \Rightarrow \frac{\sum_{i=1}^{n} X_i^2}{\theta} \sim \chi^2_{(n)}$, sehingga

$$E(Y) = E\left(\sum_{i=1}^{n} X_i^2\right) = E\left(\frac{\sum_{i=1}^{n} X_i^2}{\theta}\theta\right) = \theta E\left(\frac{\sum_{i=1}^{n} X_i^2}{\theta}\right) = n\theta$$

maka $Y_1 = \frac{\sum_{i=1}^{n} X_i^2}{n}$ adalah UUMVE untuk θ.
Akan dicari statistik Y_2 sedemikian sehingga $E(Y_2) = \theta^2$.
Setelah ditemukan, tentukan $E(Y_2|Y_1) = \varphi(Y_1)$ yang merupakan UMVE untuk θ^2.

8.2. : Informasi Fisher dan Pertidaksamaan Rao-Cramer

a. Informasi Fisher

Misal X adalah variabel random yang mempunyai p.d.f. $f(x;\theta), \theta \in \Omega$, di mana ruang parameter Ω berupa suatu interval. Informasi Fisher untuk X diberikan oleh:

$$I(\theta) = E\left[\left(\frac{\partial \ln\big(f(x;\theta)\big)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln\big(f(x;\theta)\big)}{\partial\theta^2}\right]$$

Jika $I(\theta) = 0$, maka tidak ada informasi yang diperoleh untuk θ.
Semakin besar nilai $I(\theta)$, maka semakin baik informasi yang diperoleh untuk θ.

Misal $X_1, \dots, X_n$ menyatakan sampel random dari suatu distribusi yang mempunyai p.d.f. $f(x;\theta), \theta \in \Omega$, di mana ruang parameter Ω berupa suatu interval. Informasi Fisher untuk sampel random ini diberikan oleh:

$$I_n(\theta) = nI(\theta)$$

b. Pertidaksamaan Rao-Cramer

Misal θ adalah parameter dari suatu distribusi yang ingin kita taksir.
Misal $Y = u(X_1, \dots, X_n)$ adalah penaksir untuk θ.
Definisikan $K(\theta) = E(Y)$.
Misal $Z = \sum_{i=1}^{n} \frac{\partial \ln\big(f(X_i;\theta)\big)}{\partial\theta}$. Maka $E(Z) = 0$ dan $Var(Z) = E(Z^2) = nI(\theta) = I_n(\theta)$.
Kita juga akan peroleh $K'(\theta) = E(YZ) = \rho\sigma_Y\sqrt{I_n(\theta)}$, maka diperoleh:

$$\rho = \frac{K'(\theta)}{\sigma_Y\sqrt{I_n(\theta)}}$$

Karena $\rho^2 \leq 1$, maka:

$$\frac{\big(K'(\theta)\big)^2}{\sigma_Y^2 I_n(\theta)} \leq 1 \Leftrightarrow \sigma_Y^2 \geq \frac{\big(K'(\theta)\big)^2}{I_n(\theta)}$$

Pertidaksamaan ini dinamakan pertidaksamaan Rao-Cramer.
Jika Y adalah penaksir tak bias untuk θ, maka $K(\theta) = E(Y) = \theta$ dan jika variansi dari Y adalah batas bawah dari pertidaksamaan Rao-Cramer, maka Y dikatakan penaksir yang efisien untuk θ, dengan kata lain:

$$\sigma_Y^2 = \frac{1}{I_n(\theta)} = \frac{1}{n}\frac{1}{I(\theta)} = \frac{1}{n}\frac{-1}{E\left[\frac{\partial^2 \ln(f(x;\theta))}{\partial\theta^2}\right]} = \frac{1}{n}\frac{1}{E\left[\left(\frac{\partial \ln(f(x;\theta))}{\partial\theta}\right)^2\right]}$$

c. Efisiensi Penaksir
Apabila statistik Y memiliki variansi tidak sama dengan batas bawah Rao-Cramer, maka kita bisa menghitung efisiensi statistik Y dengan membagi batas bawah Rao-Cramer dengan variansi statistik Y.
Jika statistik Y efisien, maka efisiensinya adalah 1.

9.1. : Tes Terbaik Tertentu

Pada subbab ini, kita akan mengkonstruksi uji untuk menguji hipotesis sederhana H_0 melawan hipotesis sederhana H_1, atau dengan kata lain akan ditentukan daerah kritis terbaik C. Terlebih dahulu kita definisikan notasi
$\Pr[(X_1, \dots, X_n) \in C; H_0]$ = probabilitas menolak H_0 jika H_0 benar
$\Pr[(X_1, \dots, X_n) \in C; H_1]$ = probabilitas menolak H_0 jika H_0 salah.

Definisi
Misal C himpunan bagian ruang sampel.
C disebut daerah kritis terbaik berukuran α untuk mengetes hipotesis sederhana
$H_0: \theta = \theta'$
$H_1: \theta = \theta''$
jika untuk setiap himpunan bagian dari $\mathcal{A}$ yaitu A dimana
$\Pr[(X_1, \dots, X_n) \in A; H_0] = \alpha$
berlaku
a. $\Pr[(X_1, \dots, X_n) \in C; H_0] = \alpha$ dan
b. $\Pr[(X_1, \dots, X_n) \in C; H_1] \geq \Pr[(X_1, \dots, X_n) \in A; H_1]$

Namun untuk menentukan C secara langsung dari definisi cukup sulit, sehingga kita dapat menggunakan teorema berikut.

Teorema Neyman-Pearson
Untuk suatu bilangan bulat positif n yang fixed, misal $X_1, \dots, X_n$ sampel random dari suatu distribusi dengan p.d.f $f(x;\theta)$.
P.d.f bersama (fungsi likelihood) dari $X_1, \dots, X_n$ adalah
$L(\theta; x_1, \dots, x_n) = f(x_1;\theta) \dots f(x_n;\theta)$
Misal $\Omega = \{\theta; \theta = \theta', \theta''\}$ dan k bilangan positif.
Misal C himpunan bagian dari ruang sampel sedemikian sehingga
a. $\frac{L(\theta'; x_1, \dots, x_n)}{L(\theta''; x_1, \dots, x_n)} \leq k, \forall (x_1, \dots, x_n) \in C$

b. $\frac{L(\theta';x_1,\ldots,x_n)}{L(\theta'';x_1,\ldots,x_n)} \geq k, \forall(x_1,\ldots,x_n) \in C^*$
c. $\alpha = \Pr[(X_1,\ldots,X_n) \in C; H_0]$

maka C adalah daerah kritis terbaik berukuran α untuk menguji hipotesis sederhana $H_0: \theta = \theta'$ melawan hipotesis sederhana $H_1: \theta = \theta''$.

Teorema di atas juga dapat digunakan untuk menguji hipotesis mengenai distribusi. Misal akan diuji hipotesis sederhana

$H_0: g(x_1,\ldots,x_n)$ p.d.f bersama dari $X_1,\ldots,X_n$
$H_1: h(x_1,\ldots,x_n)$ p.d.f bersama dari $X_1,\ldots,X_n$

Kedua hipotesis statistika di atas merupakan hipotesis sederhana karena secara lengkap menspesifikasikan sebuah distribusi.
Kita dapat menggunakan teorema Neyman-Pearson :
C adalah daerah kritis terbaik berukuran α untuk menguji hipotesis di atas jika untuk $k > 0$,

a. $\frac{g(x_1,\ldots,x_n)}{h(x_1,\ldots,x_n)} \leq k,, \forall(x_1,\ldots,x_n) \in C$
b. $\frac{g(x_1,\ldots,x_n)}{h(x_1,\ldots,x_n)} \geq k,, \forall(x_1,\ldots,x_n) \in C^*$
c. $\alpha = \Pr[(X_1,\ldots,X_n) \in C; H_0]$.

9.2. : Uniformly Most Powerful Tests

Pada subbab ini, kita akan mengkonstruksi uji untuk menguji hipotesis sederhana H_0 melawan hipotesis majemuk H_1.

Definisi

Daerah kritis C disebut sebagai *uniformly most powerful critical region* berukuran α untuk menguji hipotesis sederhana H_0 melawan hipotesis majemuk H_1 jika C adalah daerah kritis terbaik untuk menguji hipotesis sederhana H_0 melawan semua hipotesis sederhana di H_1.
Uji yang didefinisikan pada C disebut *uniformly most powerful test* (UMPT) dengan tingkat signifikansi α untuk menguji hipotesis sederhana H_0 melawan hipotesis majemuk H_1.

UMPT tidak selalu ada. Jika ada, kita dapat menentukannya dengan menggunakan teorema Neyman-Pearson.

Di bawah ini, akan dijelaskan hubungan UMPT dengan statistik cukup.
Misalkan $X_1,\ldots,X_n$ sampel random dari distribusi yang memiliki p.d.f $f(x;\theta), \theta \in \Omega$.
Misalkan $Y = u(X_1,\ldots,X_n)$ adalah statistik cukup untuk θ. Maka, jika ingin mengkonstruksi tes terbaik atau UMPT, maka kita tidak perlu mempertimbangkan statistik lain untuk menguji selain statistik cukup tersebut.

Jika $\theta'' < \theta'$, rasio

$$\frac{L(\theta';x_1,\ldots,x_n)}{L(\theta'';x_1,\ldots,x_n)}$$

yang bergantung pada $x_1,\ldots,x_n$ hanya melewati $y = u(x_1,\ldots,x_n)$, adalah fungsi naik dari $y = u(x_1,\ldots,x_n)$ maka dikatakan bahwa kita memiliki rasio likelihood monoton di dalam statistik $Y = u(X_1,\ldots,X_n)$.

Misalkan $X_1, \dots, X_n$ sampel random dari distribusi yang merupakan *regular case* dari kelas eksponensial,

$$f(x;\theta) = \exp[p(\theta)K(x) + S(x) + q(\theta)], \quad x \in \mathcal{A}$$
$$= 0, \qquad \text{yang lain}$$

Jika $p(\theta)$ adalah fungsi naik terhadap θ, maka kita memiliki rasio likelihood monoton di dalam statistik $Y = \sum_{i=1}^{n} K(X_i)$.

Jika kita menguji $H_0: \theta = \theta'$ melawan $H_1: \theta < \theta'$, maka, dengan $\theta'' < \theta'$,

$$\frac{L(\theta')}{L(\theta'')} \leq k$$

ekuivalen dengan $\sum_{i=1}^{n} K(x_i) \leq c$, untuk setiap $\theta'' < \theta'$. Dengan kata lain, kita memperoleh sebuah UMPT.

9.3. : Tes Rasio Likelihood

Tes rasio likelihood digunakan jika kita dihadapkan dengan masalah pengujian hipotesis majemuk H_0 melawan hipotesis majemuk H_1, atau pengujian hipotesis sederhana H_0 melawan hipotesis majemuk H_1 dimana UMPT tidak ada, contohnya $H_0: \theta = 70$ melawan $H_1: \theta \neq 70$.

Berikut langkah-langkah untuk mengkonstruksi tes rasio likelihood untuk menguji hipotesis H_0 melawan H_1 :

Misalkan $X_1, \dots, X_n$ variabel random saling bebas dengan p.d.f $f_i(x_i; \theta_1, \dots, \theta_m), i = 1, \dots, n$, dimana $(\theta_1, \dots, \theta_m) \in \Omega$.

Misal $\omega \subset \Omega$.

Akan dites :

$H_0: (\theta_1, \dots, \theta_m) \in \omega$

H_1: tidak demikian.

1. Definisikan

   $L(\omega) = f_1(x_1; \theta_1, \dots, \theta_m) \dots f_n(x_n; \theta_1, \dots, \theta_m), (\theta_1, \dots, \theta_m) \in \omega$

   $L(\Omega) = f_1(x_1; \theta_1, \dots, \theta_m) \dots f_n(x_n; \theta_1, \dots, \theta_m), (\theta_1, \dots, \theta_m) \in \Omega$
2. Cari $\widehat{\omega}$ dan $\widehat{\Omega}$ yang berturut-turut memaksimumkan $L(\omega)$ dan $L(\Omega)$.
3. Definisikan rasio likelihood

   $$\lambda(x_1, \dots, x_n) = \frac{L(\widehat{\omega})}{L(\widehat{\Omega})}$$
4. Misal $0 < \lambda_0 < 1$.

   H_0 ditolak $\Leftrightarrow \lambda(x_1, \dots, x_n) \leq \lambda_0$

   Sehingga kita mendapatkan daerah kritis

   $C = \{(x_1, \dots, x_n) \in \mathcal{A}; \lambda(x_1, \dots, x_n) \leq \lambda_0, 0 < \lambda < 1\}$

   dengan tingkat signifikansi

   $\alpha = \Pr[(X_1, \dots, X_n) \in C; H_0] = \Pr[\lambda(X_1, \dots, X_n) \leq \lambda_0; H_0]$