Bab 2

KECUKUPAN DAN KELENGKAPAN

Pre…

Misal X variabel random dengan $f(x;\theta)$ diketahui tetapi tergantung dari suatu vektor konstan berdimensi r yaitu :

$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)^t$

Ruang parameter Ω : himpunan semua nilai yang mungkin dari θ, dalam hal ini $\Omega \in \mathrm{R}^r, r \leq 1$

Misalkan $X_1, X_2, \ldots, X_n \sim iid$ dengan pdf $f(x;\theta)$

konsep kecukupan $\Rightarrow$ meringkas data tanpa menghilangkan informasi yang dibawa dalam data tentang parameter θ

Ide KECUKUPAN

- ✓ Himpunan statistik berisi banyak “informasi” dalam sampel tentang parameter, perlu adanya pembatasan statistik yang baik
- ✓ Ide kecukupan

  → pengurangan himpunan data pada statistik dengan tanpa mengurangi informasi tentang parameter

  → S dianggap sebagai statistik cukup untuk θ jika P(T|S) tidak tergantung θ, dengan T adalah statistik yang lain

- ✓ Ketika statistik cukup sudah diketahui, nilai observasi dari statistik yang lain tidak berisi informasi tentang parameter

Definisi statistik cukup

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random dari distribusi dengan parameter θ yang tidak diketahui. Statistik $Y = u(X_1, X_2, \ldots, X_n)$ disebut dengan statistik cukup untuk θ, jika distribusi bersyarat dari $X_1, X_2, \ldots, X_n$ diberikan Y tidak bergantung pada parameter θ

$$f_{X|s}(x_1, x_2, \ldots, x_n) = \begin{cases} \dfrac{f(x_1, x_2, \ldots, x_n; \theta)}{fs(s;\theta)}, & \text{jika } \sigma(x_1, x_2, \ldots, x_n; \theta) = s \\ 0, & \text{yang lainnya} \end{cases}$$

contoh

- Misalkan $X_1, X_2, \cdots, X_n$ sampel acak dari distribusi Bernoulli, dimana
  - $X_i = 1$ jika subyek ke-i menyukai sprite
  - $X_i = 0$ jika subyek ke-i tidak menyukai sprite.
- Jika p adalah peluang bahwa subyek i suka sprite, maka
  - $X_i = 1$ peluangnya p
  - $X_i = 0$ peluangnya $1 - p$.
- Misalkan sampel acak yang diambil berukuran $n = 40$ sehingga diperoleh $Y = \sum_{i=1}^{n} X_i = 22$, (di mana Y diketahui adalah banyaknya sukses dari n percobaan). Dapatkah diperoleh informasi tentang p dari sampel acak $X_1, X_2, \cdots, X_n$?

Contoh 1

Variabel random berdistribusi Bernoulli, $X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(\theta)$

Jika diketahui $\theta = P(H)$

Maka buktikan $S = \sum_{i=1}^{n} X_i$ cukup untuk θ

Contoh 2

Variabel random berdistribusi Poisson, $X_1, X_2, \ldots, X_n \sim \text{POI}(\lambda)$

Apakah $S = \sum_{i=1}^{n} X_i$ cukup untuk λ ?

Contoh 3

Variabel random berdistribusi Eksponensial, $X_i \sim \text{EXP}(\theta)$

Tentukan statistik cukup untuk θ!

Definition 10.2.1 (Bain:338)
Kriteria Faktorisasi

Jika $\mathbf{X}=\left(X_1, X_2,\ldots,X_n\right)$ dengan pdf gabungan $f\left(x_1,x_2,\ldots,x_n;\theta\right)$ dan $\mathbf{S}=\left(S_1,S_2,\ldots,S_n\right)$ adalah statistik berdimensi k maka $S_1,S_2,\ldots,S_n$ adalah himpunan statistik cukup gabungan untuk $\boldsymbol{\theta}$ jika untuk statistik lain $\mathbf{T}$, pdf gabungan T dengan syarat $S=s$, dinotasikan $f_{T|s}(t)$ tidak tergantung dari $\boldsymbol{\theta}$.

Pada kasus dimensi satu, katakan S adalah adalah statistik cukup untuk θ

Soal

1. $X_i \sim iid\ \mathrm{POI}(\lambda), \quad 0<\lambda<\infty, \lambda$ parameter tidak diketahui
2. $X_i \sim iid\ \mathrm{EXP}(\theta), \quad 0<\theta, \theta$ parameter tidak diketahui
3. Jika $X_i \sim \mathrm{iid}\ \mathrm{BIN}(1,\theta)$ maka tentukan s
4. Jika $X_i \sim \mathrm{iid}\ \mathrm{N}(\mu,\sigma^2)$ maka tentukan s
5. Jika $X_i \sim \mathrm{iid}$ Bernoulli$(p), 0\le p\le 1$. Apakah $\mathrm{T}=\sum_{i=1}^{3} X_i$ cukup untuk p?

Teorema 10.2.1 Kriteria Faktorisasi - Neyman

Jika $X = (X_1, X_2, ..., X_n)$ mempunyai pdf gabungan $f(x_1, x_2, \ldots, x_n; \theta)$

dan $S = (S_1, S_2, ..., S_n)$ maka $S_1, S_2, ..., S_n$ cukup gabungan untuk $\boldsymbol{\theta}$

jika dan hanya jika

$$f(x_1, x_2, \ldots, x_n; \boldsymbol{\theta}) = g(s; \boldsymbol{\theta}) h(x_1, x_2, \ldots, x_n)$$

$g(s; \boldsymbol{\theta})$ tidak tergantung $x_1, x_2, \ldots, x_n$ kecuali melalui s

$h(x_1, x_2, \ldots, x_n)$ tidak mengandung $\boldsymbol{\theta}$

Contoh 4.

Jika $X_i \sim$ iid BIN(1, p) maka tunjukkan bahwa $S = \sum_{i=1}^{n} X_i$ cukup untuk p

Secara umum :
Jika statistik S cukup untuk θ maka ada sebarang fungsi (1-1) S juga cukup untuk θ

Dari contoh di atas, tunjukkan bahwa $\bar{X}$ cukup untuk p!

Sifat statistik cukup

Th. 10.3.1

Jika $S_1, S_2, \ldots, S_n$ cukup gabungan untuk $\boldsymbol{\theta}$ dan $\hat{\boldsymbol{\theta}}$ MLE dari $\boldsymbol{\theta}$ maka $\hat{\boldsymbol{\theta}}$ adalah fungsi dari $S = (S_1, S_2, \ldots, S_n)$

contoh

1. Jika $X_i \sim \text{iid POI}(\lambda)$ maka tunjukkan bahwa $S = \sum_{i=1}^{n} X_i$ cukup untuk λ

2. $X_i \sim \text{iid N}(\mu, \sigma^2)$ maka tentukan S, kemudian tentukan apakah $\hat{\mu}, \hat{\sigma}^2$ cukup gabungan untuk μ, σ^2

3. $X_1, X_2 \sim iid$ pdf $f(x_1;\theta) = \theta e^{-\theta x_1}, 0 < x_1 < \infty$

$$g(x_2;\theta) = 2\theta e^{-2\theta x_2}, 0 < x_2 < \infty$$

$\theta > 0$ parameter tidak diketahui, untuk $x_1 > 0, x_2 < \infty$

tentukan statistik cukup untuk θ !

Theorema Statistik Cukup Minimal – Lehmann Scheffe

Misalkan fungsi $\mathbf{h}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}) = \dfrac{\prod_{i=1}^{n} f(x_i;\theta)}{\prod_{i=1}^{n} f(y_i;\theta)}$

adalah rasio fungsi likelihood saat $\mathbf{x}$ dan $\mathbf{y}$ dimana $\boldsymbol{\theta}$ parameter

tidak diketahui dan $\mathbf{x}$, $\mathbf{y} \in \chi^n$

Statistik $\mathbf{T} \equiv \mathbf{T}(X_1, X_2, \ldots, X_n) = (T_1, T_2, \ldots, T_k)$ yang memenuhi

untuk suatu nilai $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ dan $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ dari χ^n

nilai $\mathbf{h}(\mathbf{x},\mathbf{y};\boldsymbol{\theta})$ tidak melibatkan $\boldsymbol{\theta}$.

d.k.l

$\mathbf{T}(\mathbf{x}) = \mathbf{T}(\mathbf{y})$

$\mathbf{T}_i(\mathbf{x}) = \mathbf{T}_i(\mathbf{y}), i = 1, \ldots, k$

maka T adalah statistik cukup minimal untuk $\boldsymbol{\theta}$

Def. 10.2.2 (Bain:337)

Himpunan statistik disebut dengan cukup minimal Jika anggota himpunan merupakan cukup gabungan untuk parameter dan merupakan fungsi dari statistik cukup gabungan

contoh

$$\text{Jika } X_i \sim \text{iid Bernoulli}(p) \text{ maka tunjukkan bahwa}$$

$$S = \sum_{i=1}^{n} X_i \text{ cukup minimal untuk } p$$

Teorema

Untuk sebarang statistik yang merupakan fungsi 1-1 dari statistik cukup minimal adalah cukup minimal