Code

Tugas Praktikum 1

Angga Fathan Rofiqy

29 January, 2024

> - ***Code/Syntax*** *: File.rmd (https://github.com/Zen-Rofiqy/STA1312-PSB/blob/main/Project/Tugas%20Prak%201%20Pert%201/G1401211006_Angga-Fathan-Rofiqy_Tugas-Praktikum-1.Rmd)*

No 1

Tugas 1

Peluang Dena akan menyukai sebuah restoran adalah 0.7. Di antara restoran yang ia sukai, 20% memiliki bintang 5 di aplikasi OKfood, 50% memiliki bintang 4, dan 30% memiliki kurang dari bintang 4. Informasi apa lagi yang kita perlukan jika kita ingin mencari peluang bahwa Dena menyukai suatu restoran jika restoran tersebut memiliki kurang dari bintang 4 di aplikasi OKfood? Tunjukkan Secara Matematis mengapa informasi tersebut diperlukan?

> ***Misalkan***
>
> - D *: Dena menyukai sebuah restoran*
> - $B5$ *: Restoran rating bintang 5 di OKfood*
> - $B4$ *: Restoran rating bintang 4 di OKfood*
> - Bk *: Restoran rating bintang kurang dari 4 di OKfood*

> ***Diketahui***
>
> - $P(D) = 0.7$
> - $P(B5|D) = 0.2$
> - $P(B4|D) = 0.5$
> - $P(Bk|D) = 0.3$

> ***Ditanya:*** $P(D|Bk)$ *?*

Penyelesaian

Menggunakan **rumus bayes,**

$P(B_j|A) = \frac{P(A|B_j) \times P(B_j)}{P(A)}$, dimana $P(A) = \sum_{i=1}^{k} P(A|B_i) \times P(B_i)$

Maka,

$P(D|Bk) = \frac{P(Bk|D) \times P(D)}{P(Bk)} = \frac{0.3 \times 0.7}{P(Bk)}$

Dimana, $\begin{aligned} P(Bk) &= P(Bk|D) \times P(D) \ + \ P(Bk|D^c) \times P(D^c) \\ &= (0.3 \times 0.7) \ + \ P(Bk|D^c) \times 0.3 \end{aligned}$

Bisa dilihat bahwa informasi tambahan yang diperlukan apabila kita ingin mencari nilai $P(D|Bk)$ (Peluang Dena menyukai resto rating kurang dari 4) adalah $P(Bk|D^c)$ (Persentase resto rating kurang dari 4 yang tidak disukai Dena).

No 2

Tugas 2
Misalkan Y adalah banyakanya cacahan (*count*) dari peubah acak Poisson (μ).
Misalkan 5 kemungkinan nilai μ adalah 1, 2, 3, 4, dan 5. Kemudian misalkan saja kita tidak memiliki alasan untuk memberikan bobot yang mungkin lebih besar satu nilai dibandingkan dengan nilai yang lain, sehingga kita memberikan bobot yang sama untuk setiap nilai.
Selanjutnya diketahui bahwa $Y = 2$ telah teramati.
a. Tentukan distribusi posterior dengan mengisi tabel bayes box
b. Buatlah soal a dengan menggunakan R
c. Buatlah grafik barchart berdasarkan hasil b dengan sumbu x adalah nilai-nilai μ dan sumbu y adalah nilai peluang posterior

Poin a

Likelihood sebaran Poisson dengan $Y = 2$:

$$P(Y=2|\mu) = \frac{e^{-\mu}\mu^2}{2!} \text{ , } \mu = 1, 2, 3, 4, 5$$

Karena setiap nilai memiliki bobot yang sama, maka untuk setiap μ prior nya sama. Sehingga :

$$P(Y=2) = \frac{1}{5} \text{ , } \mu$$

> *Tabel bayes box*

μ	Prior	Likelihood	Prior $\times$ Likelihood	Posterior
1	1/5	0.1839	0.0368	0.2023
2	1/5	0.2707	0.0541	0.2976
3	1/5	0.2240	0.0448	0.2464
4	1/5	0.1465	0.0293	0.1611
5	1/5	0.0842	0.0168	0.0926

μ	Prior	Likelihood	Prior × Likelihood	Posterior
Total			0.1819	

Poin b

Hide

```
# Nilai μ yang mungkin
mu <- c(1,2,3,4,5)
k <- 2 # Y=2
w <- 0 # Bobot nya sama

# Sebaran Poisson
pois.llh <- function(k, mu) return(dpois(k, mu))

post.dist <- function(k, mu, w){
  # Prior untuk setiap nilai μ
  pri <- ifelse(w == 0, rep(1/length(mu), length(mu)), w)

  # Menghitung likelihood untuk Y=2 untuk setiap nilai μ
  llh <- sapply(mu, pois.llh, k=k)

  # Menghitung Prior x likelihood
  pri.llh <- llh * pri

  # Menghitung Posterior
  post <- pri.llh / sum(pri.llh)

  data.frame( mu, pri, llh, pri.llh, post ) %>% return()
}

bayes.box <- post.dist(k, mu, w)
colnames(bayes.box) <- c("mu", "Prior", "Likelihood", "Prior x Likelihc
bayes.box %>% datatable %>%
  formatRound(c(3,4,5), digits = 4)
```

Show 10 entries Search:

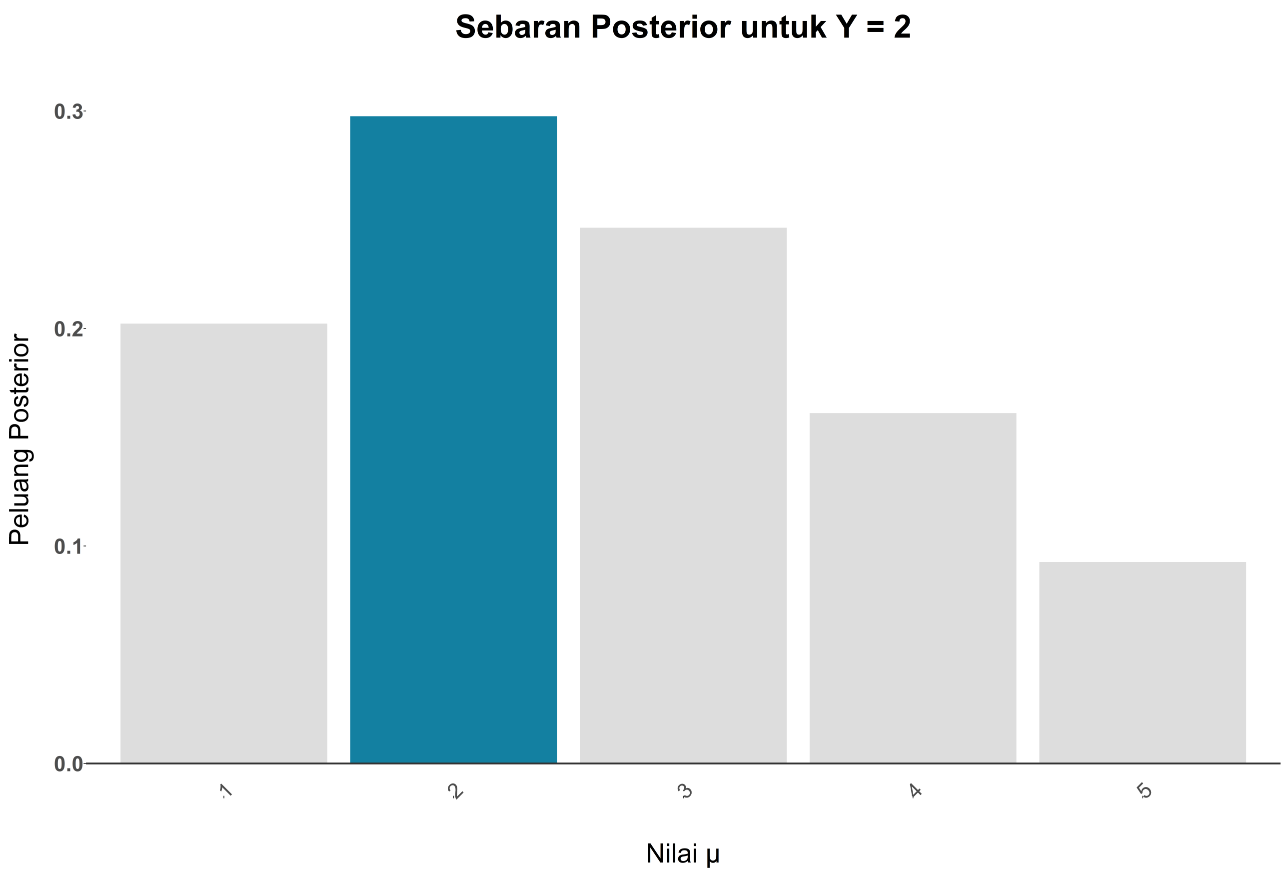
	mu	Prior	Likelihood	Prior x Likelihood	Posterior
1	1	0.2	0.1839	0.0368	0.2023
2	2	0.2	0.2707	0.0541	0.2976
3	3	0.2	0.2240	0.0448	0.2464
4	4	0.2	0.1465	0.0293	0.1611
5	5	0.2	0.0842	0.0168	0.0926

Showing 1 to 5 of 5 entries Previous 1 Next

Poin c

Hide

```
ggplot(bayes.box[,c(1,5)], aes(x=factor(mu), y=Posterior)) +
  geom_bar(stat="identity", col=NA,
           fill= ifelse(bayes.box[,c(1,5)]$mu == 2,  "#1380A1", "#ddddc
  labs(x = "\nNilai μ", y = "Peluang Posterior",
       title = "\nSebaran Posterior untuk Y = 2\n") +
  theme1.1
```



No 3

Tugas 3

Misalkan n adalah banyaknya pelanggan yang mengunjungi toko pada hari diskon dan banyaknya pelanggan tersebut tidak diketahui. Jumlah pelanggan yang melakukan pembelian adalah $Y|n \sim \text{Binomial}(n, \theta)$ di mana θ adalah peluang untuk melakukan pembelian jika pelanggan mengunjungi toko tersebut. Nilai priornya adalah $n \sim \text{Poisson}(5)$. Dengan mengasumsikan bahwa θ diketahui dan n adalah parameter yang tidak diketahui, Buat grafik dengan bantuan R distribusi posterior dari n untuk semua kombinasi $Y \in \{0,5,10\}$ dan $\theta \in \{0.2,0.5\}$ dan berikan komentar mengenai pengaruh Y dan posteriornya.

> ***Diketahui***
>
> - n *: banyaknya pelanggaan yang dateng ke toko pas hari diskon,* $n \sim Poisson(5)$
> - Y *: jumlah pelanggan yang beli,* $Y|n \sim Binomial(n, \theta)$*, dimana* $Y \in \{0, 5, 10\}$
> - θ *: peluang beli jika pelanggan dateng ke toko,* $\theta \in \{0.2, 0.5\}$

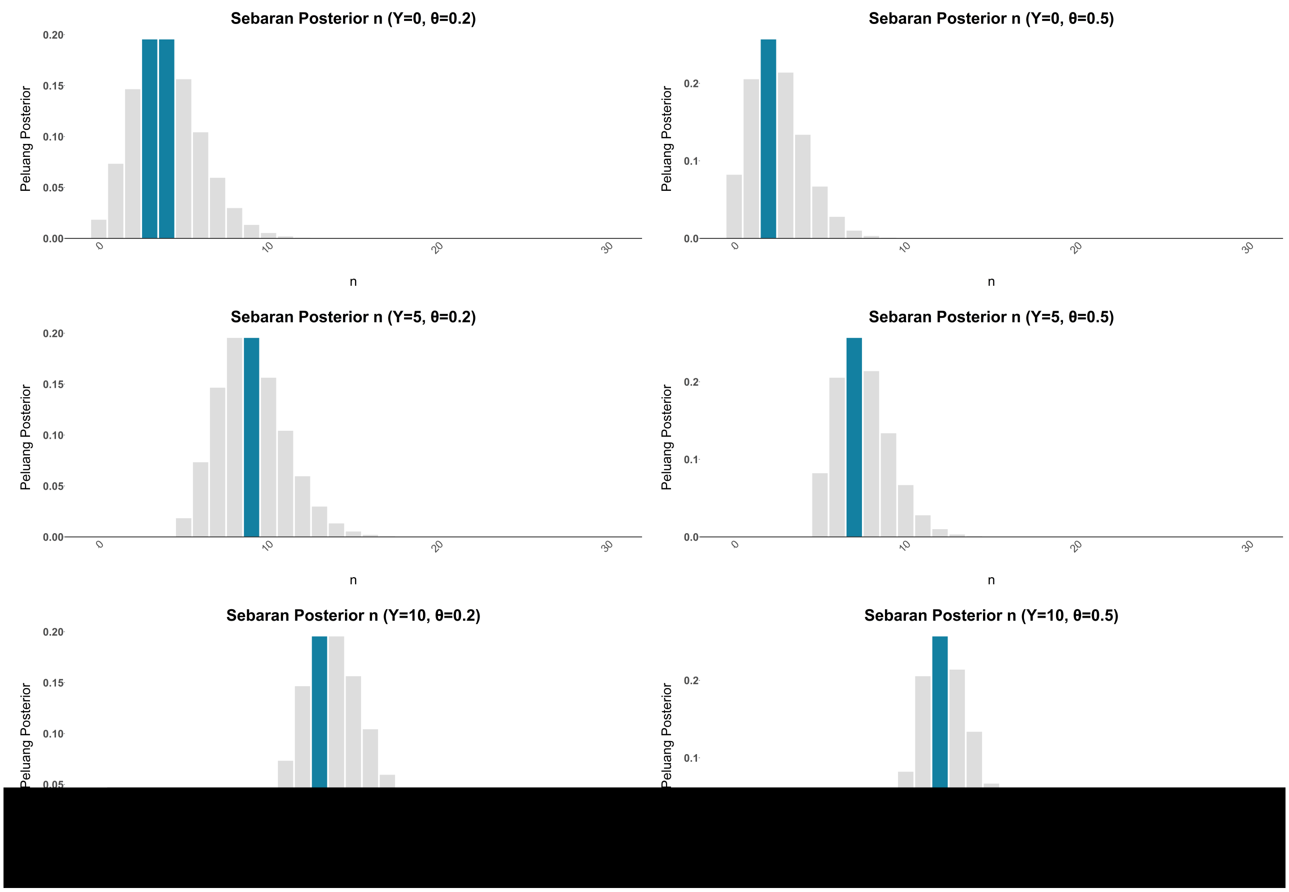
Penyelesaian

Sebaran posterior dihitung untuk setiap kombinasi Y dan θ, yakni $\binom{4}{2} = 6$, sehingga akan ada **6 grafik**.

Hide

```r
# Fungsi untuk menghitung distribusi posterior dari n
post.n <- function(y, theta, lamb.pois, n){
  # Menghitung prior Poisson
  pri <- dpois(n, lamb.pois)

  # Menghitung lilelihood Binomial P(Y|n, theta)
  llh <- dbinom(y, n, theta)

  # Menghitung posterior tidak ternomalisasi
  pri.llh <- pri * llh

  # Posterior
  post <- pri.llh / sum(pri.llh)
  return(post)
}

# Nilai-nilai yang mungkin untuk n
n.val <- 0:30

# Kombinasi nilai Y dan theta yang diberikan
Y.val <- c(0, 5, 10)
theta.val <- c(0.2, 0.5)

# Menghitung dan memplot distribusi posterior untuk semua kombinasi Y d
post.plot <- function(Y.val, theta.val, n.val, lamb.pois){
  plot.list <- list()
  for(y in Y.val){
    for(theta in theta.val){
      post <- post.n(y, theta, lamb.pois, n.val)
      df <- data.frame(n=n.val, post=post)

      #Chart
      bar.max <- which(df$post == max(df$post))
      chart <- ggplot(df, aes(x=n, y=post)) +
        geom_bar(stat="identity", col=NA,
                 fill= ifelse(df$n %in% (bar.max-1),  "#1380A1", "#dddc
        labs(x = "\nn", y = "\nPeluang Posterior",
             title = paste0("\nSebaran Posterior n (Y=",y, ", \u03b8=",
        theme1.1
      plot.list[[paste0("Y=", y, "\u03b8=", theta)]] <- chart
    }
  }
```

```
  return(plot.list)
}

# Asumsikan lambda untuk distribusi Poisson adalah 5
lamb.pois <- 5

# Plot/Chart
chart <- post.plot(Y.val, theta.val, n.val, lamb.pois)
do.call("grid.arrange", c(chart, ncol=2))
```



Dari chart di atas, terlihat bahwa nilai Y dan θ memperngaruhi sebaran posterior, yaitu:

- Ketika nilai Y **kecil**, misalkan saat $Y = 0$, sebaran posterior cenderung berepusat di nilai n yang **lebih kecil**.
  Hal ini menunjukkan jika **sangat sedikit** pelanggan yang membeli (Y), kemungkinan banyaknya pelanggan yang dateng ke toko pada hari diskon (n) juga **akan sedikit**.

- Sebaliknya, ketika nilai Y **besar** , sebaran posterior cenderung berpusat di nilai n yang **lebih besar** . Ini berarti jika **banyak** pelanggan yang membeli (Y), kemungkinan banyaknya pelanggan yang dateng ke toko pada hari diskon (n) juga **akan lebih banyak** .

- Sementara nilai θ yang **lebih besar** , misalkan saat $\theta = 0.5$ cenderung membuat nilai n yang **lebih besar** dibandingkan saat $\theta = 0.2$. Hal ini menunjukkan saat peluang untuk membeli jika pelanggan dateng ke toko (θ) **lebih tinggi**, sebaran posterior akan menunjukkan peluang yang **lebih tinggi** untuk nilai n.

Sehingga dapat disimpulkan bahwa banyaknya pelanggan yang membeli (Y) dan peluang untuk membeli jika pelanggan dateng ke toko (θ) memberikan **informasi penting** terkait banyaknya pelanggan yang dateng ke toko pada hari diskon yang tercermin dalam sebaran posterior dari n.