Kamilah Nurul Azizah
G1401211073

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:
A : Restoran yang memiliki bintang 5 di aplikasi Okfood
B : Restoran yang memiliki bintang 4 di aplikasi Okfood
C : Restoran yang memiliki kurang dari bintang 4 di aplikasi OKfood
D : Dena akan menyukai sebuah restoran

Diketahui:
$P(D) = 0.7$
$P(A|D) = 0.2$
$P(B|D) = 0.5$
$P(C|D) = 0.3$

Ditanya: $P(D|C)$?

Jawab:

$$P(D|C) = \frac{P(C|D) \times P(D)}{P(C)} = \frac{0.3 \times 0.7}{\boldsymbol{P(C)}?}$$

Untuk mencari nilai $P(D|C)$ **diperlukan nilai $\boldsymbol{P(C)}$** yaitu **peluang restoran memiliki kurang dari bintang 4 di aplikasi Okfood**.

Tugas 2

Misalkan Yadalah 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. Y adalah banyaknya cacahan dari peubah acak Poisson (μ)

- Likelihood distribusi Poisson dengan $Y = 2$

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

- Setiap nilai memiliki bobot yang sama → prior yang sama untuk setiap μ sehingga

$$P(Y = 2) = \frac{1}{5} \quad , \quad \mu = 1, 2, 3, 4, 5$$

μ	Prior	Likelihood	Prior x 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
Total			0.1819	

b. Buatlah soal a dengan menggunakan R

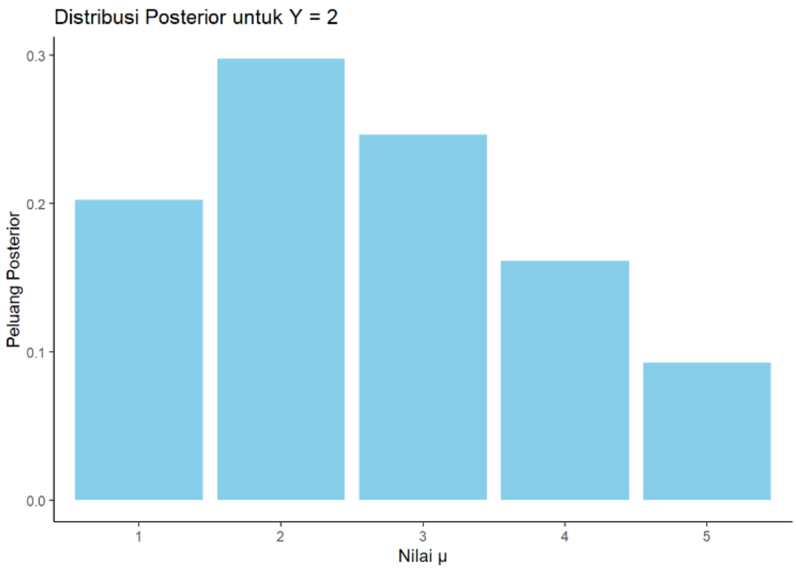
```r
# Nilai μ yang mungkin
mu <- c(1,2,3,4,5)

# Prior untuk setiap nilai μ
prior <- rep(1/length(mu), length(mu))

# Distribusi Poisson
poisson_likelihood <- function(k, mu) {
  return(dpois(k, mu))
}

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

# Menghitung Prior x Likelihood
priorxlikelihood <- likelihood * prior

# Menghitung Posterior
posterior <- priorxlikelihood / sum(priorxlikelihood)
```

```r
df <- data.frame(mu,prior,likelihood,priorxlikelihood,posterior)
knitr::kable(df, col.names = c("mu", "Prior", "Likelihood", "Prior
x Likelihood", "Posterior"), digits = c(0, 1, 4, 4, 4))
```

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

c. Buatlah grafik *barchart* berdasarkan hasil b dengan sumbu x adalah nilai-nilai μ dan sumbu y adalah nilai peluang posterior

```r
df2 <- data.frame(mu, posterior)
library(ggplot2)
ggplot(df, aes(x=factor(mu), y=posterior)) +
  geom_bar(stat="identity", fill="skyblue") +
  xlab("Nilai μ") +
  ylab("Peluang Posterior") +
  ggtitle("Distribusi Posterior untuk Y = 2") +
  theme_classic()
```



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 $Y|n$~Binomial(n, θ) di mana θ adalah peluang untuk melakukan pembelian jika pelanggan mengunjungi toko tersebut. Nilai priornya adalah n~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.

n: banyaknya pelanggan yang mengunjungi toko pada hari diskon, $n \sim Poisson(5)$

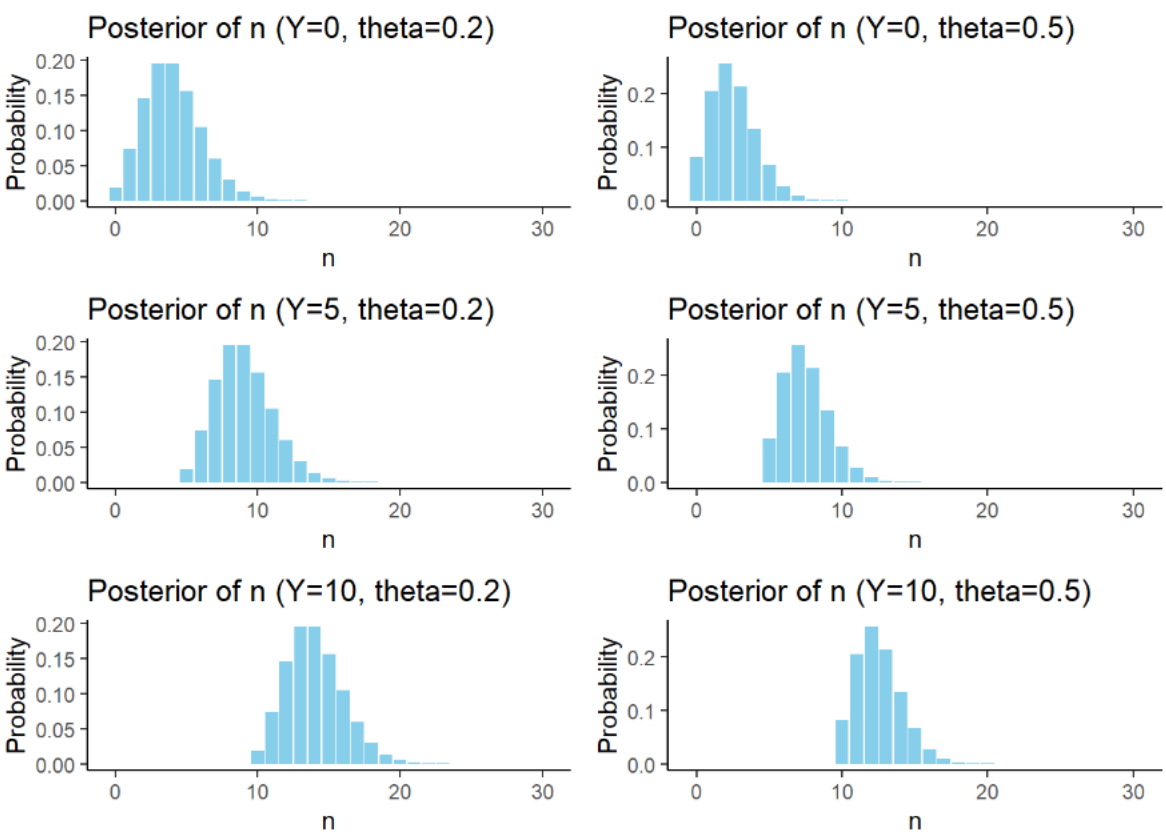
Y: jumlah pelanggan yang melakukan pembelian, $Y|n \sim Binomial(n, \theta)$, $Y \in \{0,5,10\}$

θ: peluang untuk melakukan pembelian jika pelanggan mengunjungi toko tersebut, $\theta \in \{0.2, 0.5\}$

Menghitung distribusi posterior untuk setiap kombinasi Y dan θ sehingga menghasilkan 6 grafik dengan *syntax* R sebagai berikut:

```r
library(ggplot2)

# Fungsi untuk menghitung distribusi posterior dari n
calculate_posterior <- function(y, theta, lambda_poisson, n) {
  # Menghitung prior Poisson
  prior <- dpois(n, lambda_poisson)

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

  # Menghitung posterior tidak ternormalisasi
  priorxlikelihood <- prior * likelihood

  # Posterior
  posterior <- priorxlikelihood / sum(priorxlikelihood)

  return(posterior)
}

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

# Kombinasi nilai Y dan theta yang diberikan
Y_values <- c(0, 5, 10)
theta_values <- c(0.2, 0.5)
```

```r
# Menghitung dan memplot distribusi posterior untuk semua kombinasi Y dan theta
plot_posterior <- function(Y_values, theta_values, n_values, lambda_poisson) {
  plot_list <- list()
  for (y in Y_values) {
    for (theta in theta_values) {
      posterior <- calculate_posterior(y, theta, lambda_poisson, n_values)
      df <- data.frame(n=n_values, posterior=posterior)
      p <- ggplot(df, aes(x=n, y=posterior)) +
        geom_bar(stat="identity", fill="skyblue") +
        ggtitle(paste("Posterior of n (Y=", y, ", theta=", theta, ")", sep=""))
+
        xlab("n") +
        ylab("Probability")
      plot_list[[paste("Y=", y, "theta=", theta, sep="")]] <- p
    }
  }
  return(plot_list)
}

# Asumsikan lambda untuk distribusi Poisson adalah 5
lambda_poisson <- 5

# Pplot
plot_list <- plot_posterior(Y_values, theta_values, n_values, lambda_poisson)
library(gridExtra)
do.call("grid.arrange", c(plot_list, ncol=2))
```



Dari grafik tersebut terlihat nilai Y dan θ memengaruhi distribusi posterior, yaitu:

- Ketika Y kecil misalkan saat $Y = 0$, distribusi posterior condong ke nilai n yang lebih kecil. Hal ini menunjukkan jika sangat sedikit pelanggan yang melakukan pembelian (Y), kemungkinan banyaknya pelanggan yang mengunjungi toko pada hari diskon (n) juga akan sedikit.
- Sebaliknya, ketika Y besar, distribusi posterior condong ke nilai n yang lebih besar. Hal ini menunjukkan bahwa jika banyak pelanggan yang melakukan pembelian (Y), kemungkinan banyaknya pelanggan yang mengunjungi toko pada hari diskon (n) juga lebih banyak.
- Sementara nilai θ yang lebih besar misalkan saat $\theta = 0.5$ membuat distribusi posterior lebih merata dan cenderung membuat nilai n yang lebih besar dibandingkan saat $\theta = 0.2$. Hal ini menunjukkan saat peluang untuk melakukan pembelian jika pelanggan mengunjungi toko tersebut (θ) lebih tinggi, distribusi posterior akan menunjukkan peluang yang lebih tinggi untuk nilai n.
- Sebaliknya, saat nilai θ lebih rendah, distribusi posterior akan menunjukkan peluang yang lebih tinggi untuk nilai n yang lebih besar untuk menjelaskan Y yang sama.
- Maka, dapat disimpulkan bahwa banyaknya pelanggan yang melakukan pembelian (Y) dan peluang untuk melakukan pembelian jika pelanggan mengunjungi toko tersebut (θ) memberikan informasi penting terkait banyaknya pelanggan yang mengunjungi toko pada hari diskon yang tercermin dalam distribusi posterior dari n.