Chapter 5: Bayesian Inference for Poisson Distribution

강의 목표

- 푸아송분포를 중심으로 베이지안 추론의 이해
- Parameter Estimation (모수 추정)
  - Point Estimation (점추정)
  - Confidence Interval (구간추정)
- Prediction (예측)

Poisson Distribution

- Probability mass function for Poisson with rate θ"

$$f(X = x \mid \theta) = \frac{\theta^x e^{-x}}{x!}.$$

- Suppose $x_1, ..., x_n$ have $Poi(\theta)$. Then the likelihood is

$$f(x_1, x_2, ..., x_n \mid \theta) = \prod_{i=1}^{n} f(x \mid \theta) \propto \theta^{\sum x_i} e^{-n\theta}.$$

- Sufficient Statistics: $\sum X_i$.

How to Choose Appropriate Prior

- Uniform Dist:

$$P(\theta) \propto 1.$$

- Posterior Dist:

$$P(\theta \mid x_1, \ldots, x_n) \propto \theta^{\sum x_i} e^{-n\theta}.$$

- Gamma distribution with $\sum x_i + 1$ and n.

How to Choose Appropriate Prior

- Prior Dist:

$$P(\theta) \propto \theta^{a}.$$

- Posterior Dist:

$$P(\theta \mid x_1, ..., x_n) \propto \theta^{\sum x_i + a} e^{-n\theta}.$$

- Gamma distribution with $\sum x_i + a + 1$ and n.

How to Choose Appropriate Prior

- Prior Dist:

$$P(\theta) \propto e^{-b\theta}.$$

- Posterior Dist:

$$P(\theta \mid x_1, \ldots, x_n) \propto \theta^{\sum x_i} e^{-(n+b)\theta}.$$

- Gamma distribution with $\sum x_i + 1$ and $n + b$.

How to Choose Appropriate Prior

- Prior Dist:

$$P(\theta) \propto \theta^{a+1} e^{-b\theta}.$$

- Posterior Dist:

$$P(\theta \mid x_1, \ldots, x_n) \propto \theta^{\sum x_i + a + 1} e^{-(n+b)\theta}.$$

- Gamma distribution with $\sum x_i + a$ and $n + b$.
- If the prior distribution is gamma, the posterior is gamma.

The Gamma/Poisson Bayesian Model

- Posterior Mean:

$$\hat{\lambda} = \frac{\sum x_i + a}{n + b}.$$

- It can be decomposed:

$$\hat{\lambda} = \left(\frac{n}{n+b}\right)\left(\frac{\sum x_i}{n}\right) + \left(\frac{b}{n+b}\right)\left(\frac{a}{b}\right).$$

- The data get weighted more heavily as $n \to \infty$.

Bayesian Learning

- We can use the Bayesian approach to update our information about the parameter(s) of interest **sequentially** as new data become available.
- Suppose we formulate a prior for our parameter θ and observe a random sample x_1.
- Then the posterior is

$$\pi(\theta \mid x_1) \propto p(\theta)L(\theta \mid x_1)$$

- Then we observe a new (independent) sample x_2.

Bayesian Learning

- We can use our previous posterior as the new prior and derive a new posterior:

$$\begin{aligned}\pi(\theta \mid x_1, x_2) &\propto p(\theta)L(\theta \mid x_1, x_2) \\ &\propto p(\theta)L(\theta \mid x_1)L(\theta \mid x_2) \\ &\propto p(\theta \mid x_1)L(\theta \mid x_2)\end{aligned}$$

- Note this is the same posterior we would have obtained had x_1 and x_2 arrived at the same time.
- This "sequential updating" process can continue indefinitely in the Bayesian setup.

Traffic Example

두 도시 에서 차량통행량 등 주위의 교통환경이 비슷한 교차로를 하나씩 선택하여 매주 발생한 교통사고 건수 를 1년 동안 조사하였다. 첫 번째 도시 에서는 직진 후 좌회전 신호를 사용하고 두 번째 도시 에서는 좌회전 후 직진 신호를 사용한다. 교통사고 건수는 독립적으로 포아송 분포를따 른다고 가정한다. 교통통제 등의 이유로 조사를 할 수 없었던 기간을 제외하고 다음 표와 같은 조사 결과를 얻었다.

Traffic Example

교통사고 건수	0	1	2	3	4	5	6	7
City 1의 사고 건수	7	14	13	8	4	2	2	0
City 2의 사고 건수	4	13	15	6	2	2	3	1

$$n_1 = 50, \quad \sum x_{1i} = 102, \quad \bar{x}_1 = 2.04$$

$$n_2 = 46, \quad \sum x_{2i} = 104, \quad \bar{x}_1 = 2.26$$

Traffic Example

- 두 도시의 실제 평균 교통사고 건수 θ_1과 θ_2에 대하여 $Gamma(2, 1)$의 사전분포를 가정하자.
- 그렇다면 다음과 같은 posterior distribution을 찾을 수 있다.

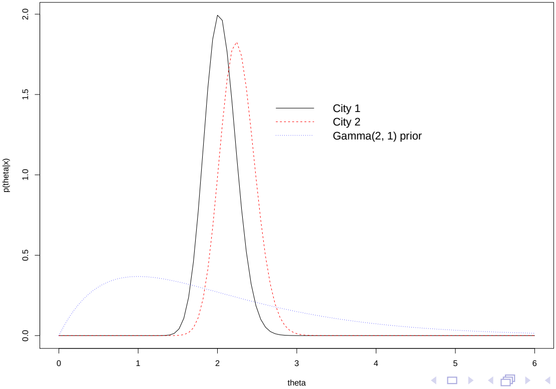
$$\pi(\theta \mid n_1 = 50, \sum x_{1i} = 102) \sim Gamma(2 + 102, 1 + 50),$$
$$\pi(\theta \mid n_1 = 46, \sum x_{1i} = 104) \sim Gamma(2 + 104, 1 + 46).$$

Traffic Example

```
x1 = rep(c(0, 1, 2, 3, 4, 5, 6), c(7, 14, 13, 8, 4, 2, 2))

x2 = rep(c(0, 1, 2, 3, 4, 5, 6, 7), c(4, 13, 15, 6, 2, 2, 3, 1))

a =2; b = 1

n1 = length(x1); s1 = sum(x1)

n2 = length(x2); s2 = sum(x2)

postmean.theta1 = (a+s1)/(b+n1)

postmean.theta2 = (a+s2)/(b+n2)

### plot the posterior

par(mfrow=c(1, 1))

theta <- seq(0, 6, length=100)

plot(theta, dgamma(theta, a+s1, b+n1), type="l", xlab="theta", ylab="p(theta|x)

lines(theta, dgamma(theta, a+s2, b+n2), lty=2, col = "red")

lines(theta, dgamma(theta, a, b), lty=3, col = "blue")

legend ( 2.5, 1.5, legend=c (paste ("City 1"), paste("City 2"),

paste("Gamma(2, 1) prior")), cex = 1.3, lty=c(1, 2, 3), col=c(1, 2, 4),

  bty="n")
```

Traffic Example

- City 1의 사고 발생 건수가 City 2에 비해 작다.
- 사후 분포의 분산이 사전 분포의 분산보다 작다.
- Likelihood의 영향으로 사후 분포들이 구간 $(1.5, 3)$이외에는 매우 비슷한다.



Prediction Distribution

- Poisson - Gamma Prediction distribution:

$$
\begin{aligned}
& f(x_{n+1} \mid x_1, \dots, x_n) \\
= & \int f(x_{n+1} \mid \theta, x_1, x_2, \dots, x_n)\pi(\theta \mid x_1, x_2, \dots, x_n) \\
= & \int f(x_{n+1} \mid \theta)\pi(\theta \mid x_1, x_2, \dots, x_n) \\
= & \int \frac{\theta^{x_{n+1}} e^{-\theta}}{x_{n+1}!} \times \frac{(b+n)^{a+\sum x_i}}{\Gamma(a+\sum x_i)} \theta^{a+\sum x_i - 1} e^{-(b+n)\theta} \\
= & \frac{(b+n)^{a+\sum x_i}}{x_{n+1}!\Gamma(a+\sum x_i)} \int \theta^{a+\sum x_i + x_{n+1} - 1} e^{-(b+n+1)\theta} \\
= & \frac{(b+n)^{a+\sum x_i}}{x_{n+1}!\Gamma(a+\sum x_i)} \times \frac{\Gamma(a+\sum x_i + x_{n+1})}{(b+n+1)^{a+\sum x_i + x_{n+1}}} \\
\propto & \frac{1}{x_{n+1}!} \times \frac{\Gamma(a+\sum x_i + x_{n+1})}{(b+n+1)^{a+\sum x_i + x_{n+1}}}
\end{aligned}
$$

Prediction Distribution

$$
\begin{aligned}
& f(x_{n+1} \mid x_1, \ldots, x_n) \\
= \; & \binom{a + \sum x_i + x_{n+1} - 1}{a + \sum x_i - 1} \left(\frac{b+n}{b+n+1} \right)^{a + \sum x_i} \left(\frac{1}{b+n+1} \right)^{x_{n+1}} \\
= \; & \binom{a + \sum x_i + x_{n+1} - 1}{x_{n+1}} \left(\frac{1}{b+n+1} \right)^{x_{n+1}} \left(\frac{b+n}{b+n+1} \right)^{a + \sum x_i} .
\end{aligned}
$$

Prediction Distribution

$$\Pr(X = x) \quad = \quad \binom{x + r - 1}{x} p^x (1 - p)^r \quad \text{for } x = 0, 1, 2, \ldots$$

Hence the prediction distribution is

$$NB\left(a + \sum_{i=1}^{n} x_i, \frac{1}{b + n + 1}\right).$$

Prediction Distribution

- 예측 기대치:

$$\mathbb{E}(X_{n+1} \mid x_1, x_2, ..., x_n) = \frac{a + \sum x_i}{b + n}.$$

- 예측기대치는 사후기대치와 동일.
- 예측 분산:

$$\mathrm{Var}(X_{n+1} \mid x_1, x_2, ..., x_n) = \frac{a + \sum x_i}{(b + n)^2}(b + n + 1).$$

- 예측분산은 사후분산보다 $(b + n + 1)$ 곱한 만큼 크다.

Prediction Expectation

- Expectation: using the fact $\mathbb{E}(X_{n+1} \mid \theta) = \theta$.

$$\begin{aligned} \mathbb{E}(X_{n+1} \mid x_1, ..., x_n) &= \mathbb{E}(\mathbb{E}(X_{n+1} \mid \theta, x_1, ..., x_n) \mid x_1, ..., x_n) \\ &= \mathbb{E}(\theta \mid x_1, ..., x_n). \end{aligned}$$

- Hence, the expected prediction is the same as the posterior expectation.

Prediction Variance

- Variance: using the fact $Var(X_{n+1} \mid \theta) = \theta$.

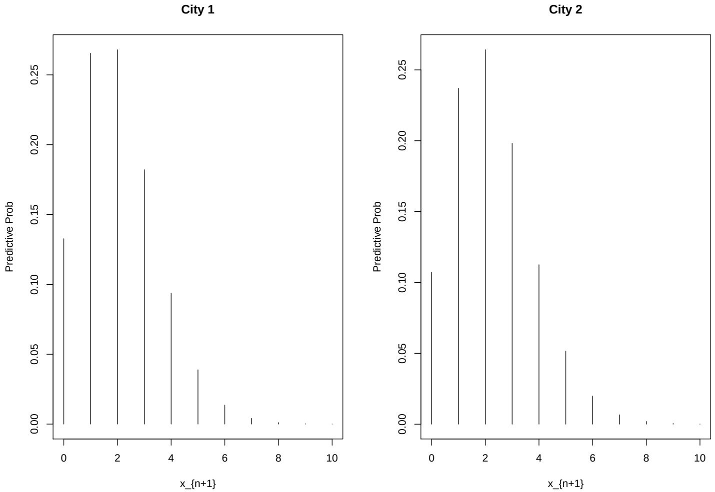
$$\begin{aligned} Var(X_{n+1} \mid x_1, ..., x_n) &= Var(\mathbb{E}(X_{n+1} \mid \theta, x_1, ..., x_n) \mid x_1, ..., x_n) \\ &\quad + \mathbb{E}(Var(X_{n+1} \mid \theta, x_1, ..., x_n) \mid x_1, ..., x_n) \\ &= Var(\theta \mid x_1, ..., x_n) + \mathbb{E}(\theta \mid x_1, ..., x_n) \end{aligned}$$

- Hence, the variance of prediction distribution is larger than the the variance of the posterior expectation.

Predictive Probability

```
> ## Ch 6
> #predictive distribution of X_{n+1}
> x1=c(rep(0,7),rep(1,14),rep(2,13),rep(3,8),rep(4,4),rep(5,2),
+      rep(6,2))
> x2=c(rep(0,4),rep(1,13),rep(2,15),rep(3,6),rep(4,2),rep(5,2),
+      rep(6,3),rep(7,1) )
> a =2;b = 1
> n1 = length(x1); s1 = sum(x1)
> n2 = length(x2); s2 = sum(x2)
> x = seq(0,10)
> par(mfrow=c(1, 2))
> plot(x,dnbinom(x,size=a+s1,prob=(b+n1)/(b+n1+1)), xlab="x_{n+1}",
+      ylab="Predictive Prob", type="h", main="City 1")
> plot(x,dnbinom(x,size=a+s2,prob=(b+n2)/(b+n2+1)), xlab="x_{n+1}",
+      ylab="Predictive Prob" ,type="h" , main="City 2" )
```

Predictive Probability



Monte Carlo Method for Traffic Example

- 두 도시의 실제 평균 교통사고 건수 θ_1과 θ_2에 대하여 다음과 같은 posterior distribution을 찾았다.

$$\pi(\theta \mid n_1 = 50, \sum x_{1i} = 102) \sim Gamma(2 + 102, 1 + 50),$$
$$\pi(\theta \mid n_2 = 46, \sum x_{2i} = 104) \sim Gamma(2 + 104, 1 + 46).$$

- Gamma 분포는 이미 알려져 있지만, 관련 statistics는 여전히 찾기 어렵다.

Monte Carlo Method for Traffic Example

- 이 예제에서 주요 목적은 θ_1과 θ_2 차이가 얼마인지 찾는데에 있다. (i.e., $\theta_1 - \theta_2$).
  - Posterior expectation for $\theta_1 - \theta_2$ given data.
  - Posterior variance for $\theta_1 - \theta_2$ given data.
- Monte Carlo Method를 통해 두 parameter들의 차이에 관련된 통계량을 찾을 수 있다

Monte Carlo Method for Traffic Example

- For notational convenience, let $\eta = \theta_1 - \theta_2$.

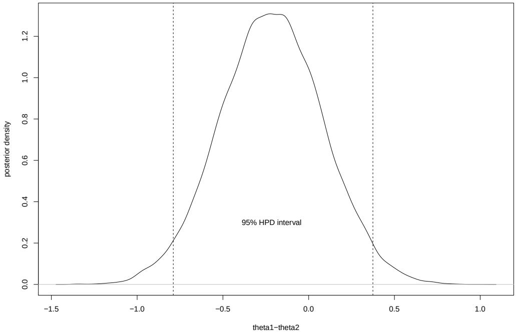
```
a =2 ; b = 1

n1 = 50; s1 = 102 ; n2 = 46; s2 = 104 ;

nsim = 30000

theta1.sim = rgamma(nsim,a+s1,b+n1)

theta2.sim = rgamma(nsim,a+s2,b+n2)

eta=theta1.sim- theta2.sim

mean(eta)

[1] -0.2155787

var( eta)

[1] 0.08880491
```

Monte Carlo Method for Traffic Example

- 이론 $\mathbb{E}(\eta \mid \text{data}) = -0.2161$.
- 이론 $(\eta \mid \text{data}) = 0.0875$.

Monte Carlo Method for Traffic Example

```
HPD=HPDsample(eta)
par(mfrow=c(1,1))
plot(density(eta), type="l", xlab= "theta1-theta2",
ylab="posterior density", main="")
abline( v= HPD, lty=2)
text(mean(eta),0.3, "95% HPD interval" )
```



Monte Carlo Method for Prediction

- 예측 분포 $X_{n+1} \mid X_1, ..., X_n$은 Posterior분포보다 더욱 복잡한 형태를 가진 경우가 많다.
- Monte Carlo Method는 $\mathbb{E}(X_{n+1} \mid X_1, ..., X_n)$과 $\mathrm{Var}(X_{n+1} \mid X_1, ..., X_n)$ 를 추정하는데 큰 도움을 준다.
- Recall that

$$f(X_{n+1} \mid X_1, ..., X_n) = \mathbb{E}(f(X_{n+1} \mid \theta) \mid X_1, ..., X_n).$$

- Here we do not discuss random sampling from the prediction distribution directly.

First Method

1. N개의 θ_i를 랜덤하게 생성.

$$\theta_i \sim \pi(\theta \mid x_1, ..., x_n).$$

2. 각 θ_i에 대하여 X_{n+1}를 생성함.

$$x_{n+1,i} \sim f(x_{n+1} \mid \theta_i).$$

- Estimation
  - i. $\widehat{\mathbb{E}}(X_{n+1} \mid X_1, ..., X_n) = \frac{1}{N} \sum_{i=1}^{N} x_{n+1,i}$
  - ii. $\widehat{\mathrm{Var}}(X_{n+1} \mid X_1, ..., X_n) = \frac{1}{N} \sum_{i=1}^{N} x_{n+1,i}^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_{n+1,i}\right)^2$

Second Method: Rao-Blackwellization

1. N개의 θ_i를 랜덤하게 생성.

$$\theta_i \sim \pi(\theta \mid x_1, ..., x_n).$$

2. θ_i를 통해 prediction mass function을 예측.

$$\hat{f}(x_{n+1} \mid x_1, ..., x_n) = \frac{1}{N} \sum_{i=1}^{N} f(x_{n+1} \mid \theta)$$

- Estimation
  i. $\widehat{\mathbb{E}}(X_{n+1} \mid X_1, ..., X_n) = \sum_{\text{all } x_{n+1}} x_{n+1} \hat{f}(x_{n+1} \mid x_1, ..., x_n)$
  ii. $\mathrm{Var}(X_{n+1} \mid X_1, ..., X_n) = \sum_{\text{all } x_{n+1}} x_{n+1}^2 \hat{f}(x_{n+1} \mid x_1, ..., x_n) - \left(\sum_{\text{all } x_{n+1}} x_{n+1} \hat{f}(x_{n+1} \mid x_1, ..., x_n)\right)^2.$

Second Method: Rao-Blackwellization

- Rao-Blackwellizationsm는 probability mass function을 추측.
  또 그 확률을 통해 mean과 variance를 예측.
- 첫번째 방법이 더 복잡하나, 더 정확한 경향이 있다.