Handout

Bayesian Inference

Given a prior distribution $\pi(\theta)$ and a model for some observations $f(x|\theta) = f(x_1, x_2, x_3, \ldots x_n|\theta)$ the posterior distributions is given by

Posterior distribution
$$\boxed{\pi(\theta|\boldsymbol{x}) = \pi(\theta|\boldsymbol{X} = \boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{m(\boldsymbol{x})}.}$$

Where

Marginal distribution of data

$$m(\boldsymbol{x}) = \sum_{\theta} f(x|\theta)\pi(\theta)$$
for discrete prior distributions π

$$m(\boldsymbol{x}) = \int_{\theta} f(x|\theta)\pi(\theta)d\theta$$
for continuous prior distributions π

This holds true for pmf and for pdfs.

Conjugate families for Bayesian inference

Model $f(\boldsymbol{x}\|\theta)$	Prior $\pi(\theta)$	Posterior $\pi(\theta\|\boldsymbol{x})$
Poisson(θ)	Gamma$(\alpha,\ \lambda)$	Gamma$(\alpha + n\bar{X},\ \lambda + n)$
Binomial(k, θ)	Beta$(\alpha,\ \beta)$	Beta$(\alpha + n\bar{X},\ \beta + n(k - \bar{X}))$
Normal(θ, σ)	Normal(μ, τ)	Normal$\left(\frac{n\bar{X}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2}, \frac{1}{\sqrt{n/\sigma^2 + 1/\tau^2}}\right)$

Bayesian Estimate

$$\hat{\theta}_{\mathrm{B}} = \mathbf{E}\{\theta|\boldsymbol{X} = \boldsymbol{x}\} = \begin{cases} \sum_{\theta} \theta\pi(\theta|\boldsymbol{x}) \\ \int_{\theta} \theta\pi(\theta|\boldsymbol{x})d\theta \end{cases}$$
depending on discrete or continuous posterior

The variance gives posterior risk

$$\rho(\hat{\theta}) = \mathrm{Var}\{\theta|\boldsymbol{x}\}$$

Bayesian Credible set

DEFINITION 10.5

Set C is a $(1-\alpha)100\%$ **credible set** for the parameter θ if the posterior probability for θ to belong to C equals $(1-\alpha)$. That is,

$$\boldsymbol{P}\{\theta \in C \mid \boldsymbol{X}=\boldsymbol{x}\} = \int_C \pi(\theta|\boldsymbol{x})d\theta = 1-\alpha.$$

If the posterior is Normal (or can be approximated as Normal), This is given by.

$$\mu_x \pm z_{\alpha/2}\tau_x = [\mu_x - z_{\alpha/2}\tau_x, \mu_x + z_{\alpha/2}\tau_x]$$

Simulating sampling of Random Variable on the basis of samples from U(0,1)

- Basic distributions:
  - Bernoulli(p)
    1) If $u < p$ return 1 else return 0
  - Binomial(n, p)
    1) Generate $u_1, u_2 \dots u_n$
    2) Count the number of $u_i < p$
  - Geometric(p)
    1) Keep generating $u_1, u_2 \dots$ in sequence till $u_i < p$
    2) Return i
  - Negative-Binomial(k, p)
    1) Generate k samples from Geometric(p)
    2) Add the values together
- Discrete distributions
  - Method 1
    1) Generate u
    2) Find i such that $F(i-1) \le u < F(i)$, where F(x) is the cumulative distribution function
  - Method 2
    1) Generate u
    2) Find the smallest possible value of i such that $F(i) > u$, where F(x) is the cumulative distribution function
- Continuous distributions
  - Method 1 (Rejection Method)
    1) Find a, b, x such that a,b and 0,c forms a bounding box on f(x) where f(x) is the probability distribution function [for all $a \le x \le b$, $0 \le f(x) \le c$]
    2) Generate u_1, u_2
    3) $X = a + (b-a)u_1$ and $Y = cu_2$
    4) If $Y \le f(X)$ accept X as the desired sample. Else return to step 2

- Method 2 (Inverse Method)
  1) Generate u
  2) Return $F^{-1}(u)$ where $F^{-1}(x)$ is the inverse of F(x), the cumulative density function

♦ Special Methods

- Uniform(a, b)
  1) Generate u
  2) Return u * (b-a) + a
- Poisson (λ)
  1) Generate u_1, u_2 …
  2) Find the largest value k for which $u_1 * u_2 * \ldots * u_k >= e^{-\lambda}$
  3) Return k
- Normal(μ, σ) [Box-Mueller Transform]
  1) Generate u_1, u_2
  2) $z_1 = \sqrt{-2\ln(u_1)}\cos(2\pi u_2)$
  3) $z_2 = \sqrt{-2\ln(u_1)}\sin(2\pi u_2)$
  4) $x_1 = z_1\,\sigma - \mu$
  5) $x_2 = z_2\,\sigma - \mu$

Monte Carlo Methods

Represent any complex distribution in terms of simpler distributions and use the given methods to generate samples

Markov Process

Transition probability matrix (P) gives probability of going from state to state in one step

h-step transition probability matrix

$$\boxed{P^{(h)} = \underbrace{P \cdot P \cdot \ldots \cdot P}_{h \text{ times}} = P^h}$$

If P_0 is probability distribution over possible states at time 0, forecast distribution at time h is given by

Distribution of $X(h)$

$$\boxed{P_h = P_0 P^h}$$

Long term forecast is given by steady state distribution π.

Steady-state distribution

$$\pi = \lim_{h\to\infty} P_h$$
is computed as a solution of
$$\begin{cases} \pi P &= \pi \\ \sum_x \pi_x &= 1 \end{cases}$$

Only regular markov chains or irregular markov chains with absorbing states or zones will have steady state distributions.

Counting Process

- Binomial Counting process

$$\begin{array}{lcl} \lambda & = & \text{arrival rate} \\ \Delta & = & \text{frame size} \\ p & = & \text{probability of arrival (success)} \\ & & \quad \text{during one frame (trial)} \\ X(t/\Delta) & = & \text{number of arrivals by the time } t \\ T & = & \text{interarrival time} \end{array}$$

Binomial counting process

$$\begin{array}{lcl} \lambda & = & p/\Delta \\ n & = & t/\Delta \\ X(n) & = & Binomial(n, p) \\ Y & = & Geometric(p) \\ T & = & Y\Delta \end{array}$$

- Poisson Process

Poisson process

$$\begin{array}{lcl} X(t) & = & Poisson(\lambda t) \\ T & = & Exponential(\lambda) \\ T_k & = & Gamma(k, \lambda) \\ \boldsymbol{P}\{T_k \le t\} & = & \boldsymbol{P}\{X(t) \ge k\} \\ \boldsymbol{P}\{T_k > t\} & = & \boldsymbol{P}\{X(t) < k\} \end{array}$$

Queuing Process

$$
\begin{array}{lcl}
\multicolumn{3}{c}{\underline{\text{Parameters of a queuing system}}} \\
\lambda_A & = & \text{arrival rate} \\
\lambda_S & = & \text{service rate} \\
\mu_A & = & 1/\lambda_A = \text{mean interarrival time} \\
\mu_S & = & 1/\lambda_S = \text{mean service time} \\
r & = & \lambda_A/\lambda_S = \mu_S/\mu_A = \text{utilization, or arrival-to-service ratio}
\end{array}
$$

$$
\begin{array}{lcl}
\multicolumn{3}{c}{\underline{\text{Random variables of a queuing system}}} \\
X_s(t) & = & \text{number of jobs receiving service at time } t \\
X_w(t) & = & \text{number of jobs waiting in a queue at time } t \\
X(t) & = & X_s(t) + X_w(t), \\
 & & \text{the total number of jobs in the system at time } t \\
S_k & = & \text{service time of the } k\text{-th job} \\
W_k & = & \text{waiting time of the } k\text{-th job} \\
R_k & = & S_k + W_k, \text{ response time, the total time a job spends in the} \\
 & & \text{system from its arrival until the departure}
\end{array}
$$

Little's Law (Applies to all queuing process)

$$\lambda_A \, \mathbf{E}(R) = \mathbf{E}(X)$$

$$\mathbf{E}(X_w) = \lambda_A \, \mathbf{E}(W)$$

$$\mathbf{E}(X_s) = \lambda_A \, \mathbf{E}(S) = \lambda_A \mu_S = r$$

Transition probabilty of a Bernoulli single server queing process (Infinite Capacity)

$$
\begin{array}{lcll}
p_{00} & = & \boldsymbol{P}\{\text{ no arrivals }\} & = 1 - p_A \\
p_{01} & = & \boldsymbol{P}\{\text{ new arrival }\} & = p_A \\
p_{i,i-1} & = & \boldsymbol{P}\{\text{ no arrivals } \bigcap \text{ one departure }\} & = (1-p_A)p_S \\
p_{i,i} & = & \boldsymbol{P}\{\text{ no arrivals } \bigcap \text{ no departures }\} & \\
 & & + \boldsymbol{P}\{\text{ one arrival } \bigcap \text{ one departure }\} & = (1-p_A)(1-p_S) + p_A p_S \\
p_{i,i+1} & = & \boldsymbol{P}\{\text{ one arrival } \bigcap \text{ no departures }\} & = p_A(1-p_S)
\end{array}
$$

Transition probabilty of a Bernoulli single server queing process (Capacity C)

$$
\begin{array}{lllll}
p_{00} & = & \boldsymbol{P}\{\text{ no arrivals }\} & = & 1-p_A \\
p_{01} & = & \boldsymbol{P}\{\text{ new arrival }\} & = & p_A \\
p_{i,i-1} & = & \boldsymbol{P}\{\text{ no arrivals } \bigcap \text{ one departure }\} & = & (1-p_A)p_S \\
p_{i,i} & = & \boldsymbol{P}\{\text{ no arrivals } \bigcap \text{ no departures }\} & & \\
 & & +\ \boldsymbol{P}\{\text{ one arrival } \bigcap \text{ one departure }\} & = & (1-p_A)(1-p_S)+p_Ap_S \\
p_{i,i+1} & = & \boldsymbol{P}\{\text{ one arrival } \bigcap \text{ no departures }\} & = & p_A(1-p_S)
\end{array}
$$

$$p_{C,C-1} = (1-p_A)p_S$$

$$p_{C,C} = (1-p_A)(1-p_S) + p_Ap_S + p_A(1-p_S) = 1-(1-p_A)p_S$$

Properties and Performace of a M/M/1 Process

$$\pi_x = \boldsymbol{P}\{X=x\} = r^x(1-r) \quad \text{for } x = 0,1,2,\ldots$$

$$\mathbf{E}(X) = \frac{r}{1-r}$$

$$\mathrm{Var}(X) = \frac{r}{(1-r)^2}$$

$$\text{where } r = \lambda_A/\lambda_S = \mu_S/\mu_A$$

$$
\begin{array}{lllll}
\mathbf{E}(R) & = & \dfrac{\mu_S}{1-r} & = & \dfrac{1}{\lambda_S(1-r)} \\
\mathbf{E}(W) & = & \dfrac{\mu_S\, r}{1-r} & = & \dfrac{r}{\lambda_S(1-r)} \\
\mathbf{E}(X) & = & \dfrac{r}{1-r} & & \\
\mathbf{E}(X_w) & = & \dfrac{r^2}{1-r} & &
\end{array}
$$

$$
\begin{array}{lll}
\boldsymbol{P}\{\text{server is busy}\} & = & r \\
\boldsymbol{P}\{\text{server is idle}\} & = & 1-r
\end{array}
$$

Binomial Distribution

- It is used to model number of success in a sequence of **independent** Bernoulli trials
- Models the probability of x successes in n trials
- p = probability of success; n = number of trials

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$E(X) = np$$
$$Var(X) = np(1-p)$$

4

Geometric Distribution

- It is used to model number trials needed to achieve the first success in a sequence of **independent** Bernoulli trials
- Models the probability of x^{th} successive trial resulting in a success
- p = probability of success

$$P(x) = (1-p)^{x-1} p$$
$$E(X) = \frac{1}{p}$$
$$Var(X) = \frac{1-p}{p^2}$$

5

Poisson Distribution

- It is used to model number rare events occurring within a fixed period of time
- Models the probability of x rare events occurring in a fixed period of time if we know the frequency at which the events occur on average
- λ = frequency (average number of events in a fixed time period)

$$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$
$$E(X) = \lambda$$
$$Var(X) = \lambda$$

Side Note: Poisson Approx. of Binomial

- If the number of trials is large and the probability of success is low, then we can use Poisson Distribution to approximate the Binomial Distribution
  - Also works if probability of failure is very low
- $\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} \binom{n}{x} p^x (1-p)^x = e^{-\lambda}\frac{\lambda^x}{x!}$
- Can use this approximation if n >= 30 and p <= 0.05

Gamma Distribution

- Used to model total time of multistage processes with α steps (shape parameter) where time of each step can be modeled as a Exponential distribution with frequency λ.

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda} \quad if\ \alpha > 0\ x > 0$$

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad if\ \alpha > 0$$

$$\text{also } \Gamma(\alpha) = (n-1)! \quad if\ \alpha\ is\ a\ positve\ integer$$

$$E(X) = \frac{\alpha}{\lambda}$$

$$Var(X) = \frac{\alpha}{\lambda^2}$$

- Please note that Gamma(1, λ) = Exponential(λ)

Side-Note: Gamma-Poisson Formula

- Can be used to simplify calculation of probabilities of RV T with Gamma Distribution.

$$\{T > t\} = \{X < \alpha\}$$

- Where is T has Gamma distribution with parameters α (number of events) and λ (frequency of each event).
- X models the number of events that occurs before time t. It has Poisson distribution with parameter λt.

  So,

$$P\{T > t\} = P\{X < \alpha\}$$
$$P\{T \le t\} = P\{X \ge \alpha\}$$

Where T has Gamma(α, λ) distribution and X has Poisson(λt) distribution

Normal Distribution

- Used to model a large number of scenarios
  - Sums, averages or errors: Mainly due to CLT
  - Naturally occurring phenomena
- Allows you to model a scenario on the basis of expectation μ (location parameter) and standard deviation σ (scale parameter)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} - \infty < x < \infty$$
$$E(X) = \mu$$
$$Var(X) = \sigma^2$$

Side-Note: Using Normal to Approx. Binomial

- A binomial distribution is a sum of n Bernoulli trials. So if n is large (>=30) but p is not small enough (or large enough) to use Poisson approximation (0.05 <= p <= 0.95) then we can model the binomial as a sum of Bernoulli distributions with mean p and variance p(1 - p).
- So by Central Limit Theorem,

$$\text{Binomial}(n, p) \approx Normal\left(\mu = np, \sigma = \sqrt{np(1-p)}\right)$$

- This normal distribution can be calculated by converting it to a standard normal distribution