S426/S26 Bayesian Theory and Data Analysis

General points about Bayesian methods

- Logical conclusions based on the *laws of probability*.
- Incorporates *expert* opinion in the form of *prior* information.
- Prior information should be independent of the *data* and quantified with probability.
- Combine data and prior knowledge with probability rules.
- Predictions are fundamental. Most areas of science and technology are about prediction.
- The basis of Bayesian inference is *conditional probability*.

Examples in Chapter 1 of CJBH book

- **Example 1.1:** Production process θ proportion of defective parts. 2430 parts are examined.
- Y= no. of defective parts. $Y|\theta \sim \mathit{Binomial}(2430, \theta)$ (sampling distribution).
- 'Vice-President' thinks that $P(\theta \leq 0.05) = .025$, $P(\theta \geq 0.15) = 0.025$.
- If $\theta \sim \mathit{Beta}(a, b)$ then $\theta \sim \mathit{Beta}(12.05, 116.06)$ is a prior.
- *Goal:* Obtain *posterior distribution* with Bayes theorem.
- Will check that,

$$\theta|y \sim \mathit{Beta}(a + y, n - y + b).$$

- If $y = 219$, $n = 2430$, $\theta|y = 219 \sim \mathit{Beta}(231.05, 2327.06)$.

- **Example 1.2:** y_i zinc percentage of sample i, $i = 1, 2, \ldots, 12$.
- $y_1, \ldots, y_{12} | \mu, \sigma^2 \sim N(\mu, \sigma^2)$.
- *iid* = independent and identically distributed.
- Need prior on (μ, σ^2).
- A VP of operations outlines: with 95% probability the mean percentage is between 4.5 and 5 centered at 4.75.
- Therefore, $E(\mu) = 4.75$ and $P(4.5 < \mu < 5) = 0.95$.
- We can assume μ follows a Normal distribution, $\mu \sim N(m, v)$.
- Set $m = 4.75$ and $(1.96)(v)^{1/2} = 0.25$. Then $v = 0.0163$.

- Prior on σ^2? **Reference prior:** 'common base for people to evaluate data'.
- Work with *precision* $1/\sigma^2$.
- $\sigma^{-2} \sim Gamma(0.001, 0.001)$.
- Prior mean =1. Variance = $.001/(0.001)^2$ = 1000.
- What is $Pr(a < \mu < b | y_1, y_2, \ldots, y_{12})$?
- What is the value $\tilde{\mu}$ so that $Pr(\mu < \tilde{\mu} | y_1, y_2, \ldots, y_{12}) = 0.5$?
- If y_{13} is the zinc value of a future batch, one may want to find,

$$P(Y_{13} \leq 4.4 | y_1, y_2, \ldots, y_{12})$$

(posterior predictive probability.)

- Naive solution: Find point estimators $\hat{\mu}$ and $\hat{\sigma^2}$.
- Use a Normal distribution $P(Y_{13} \leq 4.4|\hat{\mu}, \hat{\sigma^2}) = \Phi\left(\frac{4.4-\hat{\mu}}{\hat{\sigma}}\right)$.
- $\Phi(\cdot)$ is the cumulative distribution function of a standard normal.
- Treats estimates as *true* parameters values.
- Attempt to compute,

$$\int\int P(y_{13} \leq 4.4|\mu, \sigma^2)\pi(\mu, \sigma^2|y_1, y_2, \ldots, y_{12})d\mu d\sigma^2$$

- Sorry, Bayesian answers require solving multidimensional integrals!

- **Example 1.3:** $n = 38$ of *Ache* men (tribe in Paraguay).
- $y_i =$ no. of armadillos killed by the i-th man in a given day.
- $y_1, y_2, \ldots, y_n | \theta \sim Poisson(\theta)$ so θ is the 'kill rate'.
- Dr. Mc. Millan an 'expert' believes that $\theta \approx 0.5$ which we take as the median of the prior.
- 95 % probability that the *mean daily* number of kills is no greater than 2.
- $\theta \sim Gamma(a, b)$. Find *a* and *b* so that

$$P(\theta \leq 0.5 | a, b) = 0.5, P(\theta \leq 2 | a, b) = 0.95.$$

- The prior is $\theta \sim Gamma(1.11, 1.64)$.

- The posterior distribution for θ is

$$\theta | y_1, y_2, \ldots, y_n \sim Gamma\left(\sum_{i=1}^{n} y_i + a, n + b\right)$$

- For the Ache data, $n = 38$, $\sum_{i=1}^{38} y_i = 10$

$$\theta | y_1, y_2, \ldots, y_n \sim Gamma(11.1, 39.61)$$

- Posterior median, $p(\theta \leq 0.5 | y_1, y_2, \ldots, y_{38}) = 0.272$ and a 95% probability interval is $(0.140, 0.468)$.

- **Example 1.4:** Two groups of cows "infected", "uninfected".
- Interest on time to abortion.
- Group 1: $y_{1,1}, y_{1,2}, \ldots, y_{1,19} \sim LN(\mu_1, \sigma_1^2)$.
- Group 2: $y_{2,1}, y_{2,2}, \ldots, y_{2,26} \sim LN(\mu_2, \sigma_2^2)$.
- If $Y \sim LN(\mu, \sigma^2)$ then $log(Y) \sim N(\mu, \sigma^2)$.
- Apparently no expert opinion on model parameters.
- *Reference priors:* $\mu_i \sim N(0, 1000)$ and $1/\sigma_i^2 \sim Gamma(0.001, 0.001)$, $i = 1, 2$.
- Median times, $exp(\mu_1), exp(\mu_2)$.
- Parameter of interest: $\Delta = exp(\mu_1) - exp(\mu_2)$.

- Posterior distribution:
  $\Delta | y_{1,1}, y_{1,2}, \ldots, y_{1,19}, y_{2,1}, y_{2,2}, \ldots, y_{2,26}$.
- 95 % probability interval for Δ is $(1.7, 55.2)$.
- Provides evidence that time to abortion for Group 1 is greater than for Group 2.
- Since data is assumed to have a log normal distribution, is not sufficient to provide an interval for $\mu_1 - \mu_2$.
- May find a point estimator,

$$\hat{\Delta} = exp(\hat{\mu_1}) - exp(\hat{\mu_2})$$

  with $\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} log(Y_{ij}), i = 1, 2$.
- Confidence interval for Δ?

- **Example 1.5:** Y_{ij} $i = 1, \ldots, 38, j = 1, \ldots, n_i$ daily number of armadillos killed by 38 males of an Ache tribe over several *forest treks*.
- Interest to model how a man's age affects daily kill success.
- A *sampling* model can be specified as

$$
\begin{aligned}
y_{ij}|\lambda_i &\sim Poisson(\lambda_i), i = 1, \ldots, 38; j = 1, \ldots, n_i \\
log(\lambda_i) &= \beta_1 + \beta_2(a_i - \bar{a}) + \beta_3(a_i - \bar{a})^2 + \delta_i \\
\delta_i|\tau &\sim N(0, \tau^{-1})
\end{aligned}
$$

- λ_i is the kill rate and a_i is the age for subject i.
- $\bar{a}$ is the average age.

- Model assumes a quadratic effect of the log of kill rate and age.
- δ_i is a *random effect* and measures the natural ability.
- τ is the precision of the random effects.
- Due to lack of prior information, $\beta_i \sim N(0, 1000), i = 1, 2, 3$
- $\tau \sim Gamma(0.001, 0.001)$.
- An estimate of the quadratic term β_3 with an interval can provide if the effect is non-zero.
- Also,

$$\hat{\lambda}(a) = \hat{\beta}_1 + \hat{\beta}_2(a - \bar{a}) + \hat{\beta}_3(a - \bar{a})^2$$

Conditional probability and Bayes theorem

- For two events A, B, the conditional probability of A given B is
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
assuming that $P(B) > 0$.
- Also,
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$
- *Total probability*, $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$.
- *Bayes theorem*,
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

- **Drug Screening Example:** 'D' indicates *drug user*. 'C' someone who is clean.
- "+" test result is positive. "-" test result is negative.
- Overall prevalence of "D" is established by $P(D) = 0.01$. Drug use is rare.
- $P(+|D) = 0.98$ "sensitivity of the test". $P(-|C) = 0.95$ "specificity of the test".
- Applying Bayes theorem,

$$
\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|C)P(C)} \\
&= \frac{(0.98)(0.01)}{(0.98)(0.01) + (0.05)(0.99)} = 0.165
\end{aligned}
$$

- $P(D)$ is the *prior probability*. $P(D|+)$ is the *posterior probability*.
- Conditional on the test giving a "+", the posterior probability is more than 16 times greater than the prior.
- If $P(D) = 0.5$ (50/50 chance of drug use),

$$P(D|+) = \frac{(0.98)(0.5)}{(0.98)(0.5) + (0.05)(0.95)} = 0.95$$

  increased by a factor of 1.9.
- $P(D|+)/P(D)$ is a *posterior* to *prior* ratio.