Bayes theorem. Statistical framework.

- θ is some unknown **parameter**.
- θ is typically a vector, $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$.
- $r = 1$ is the scalar case or one parameter model.
- $y = (y_1, y_2, \ldots, y_n)$ is a **data** vector. Provides information about θ.
- $f(y|\theta)$ is the **sampling distribution** of y given θ.
- $p(\theta)$ is the **prior distribution** on θ. Represents our degree of belief on θ.
- θ is treated as a **random** variable.

- **Posterior distribution:** $p(\theta|y)$ is the updated knowledge about θ conditional on y.
- **Bayes theorem:**

$$p(\theta|y) \propto f(y|\theta)p(\theta)$$

- The complete formulation is:

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int_{\Theta} f(y|\theta)p(\theta)d\theta}.$$

Inference on Binomial model

- y number of successes over n trials; $y = 0, 1, 2, \ldots, n$.
- n is a fixed known quantity.
- θ is the probability of success for a single trial; $0 < \theta < 1$.
- **Binomial distribution:**

$$f(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

- A possible prior on θ, $\theta \sim U(0,1)$ or $p(\theta) = 1; 0 < \theta < 1$.
- Posterior distribution,

$$p(\theta|y) \propto \binom{n}{y} \theta^y (1-\theta)^{n-y} \propto \theta^{y+1-1}(1-\theta)^{n-y+1-1}$$

- This defines a Beta density, $Beta(y+1, n-y+1)$.
- A Bayes estimate of θ could be $\frac{y+1}{n-y+1}$, the mean of the distribution.
- Since $\theta|y \sim Beta(y+1, n-y+1)$, then $E(\theta|y) = \frac{y+1}{n-y+1}$.
- In classical methods, the same probability is estimated as $\frac{y}{n}$.
- If n is *large*, the estimators are similar.
- Bayes argued for this prior based on the fact that *prior predictive density*, $p(y) = \frac{1}{n+1}; y = 0, 1, \ldots, n$.
- Laplace considered the *Principle of insufficient reason*, "all values of θ are equally likely".

Prior predictive distribution for y

- Suppose $y_i \sim Bernoulli(\theta)$; $y_i = 0, y_i = 1$ and $\theta \sim U(0,1)$.
- Prior predictive probability for $y_i = 1$,

$$P(y_i = 1) = \int_0^1 Pr(y_i = 1|\theta)p(\theta)d\theta = \int_0^1 \theta d\theta = 1/2$$

- If y follows a full $Binomial(n, \theta)$ observation,

$$\begin{aligned} p(y) &= \int_0^1 f(y|\theta)p(\theta)d\theta = \int_0^1 \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta \\ &= \binom{n}{y} \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} = \frac{1}{n+1}. \end{aligned}$$

- This is a *discrete uniform* distribution on $\{0, 1, 2, \ldots, n\}$.
- Why? For any *Beta*(a, b) distribution,

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a-1}(1-\theta)^{b-1} d\theta = 1$$

- If z is a positive integer, $\Gamma(z) = (z-1)!$
- Apply results for $a = y+1, b = n+1$.
- If $\tilde{y}$ is a future Bernoulli outcome given that we had observed y successes,

$$\begin{aligned} P(\tilde{y} = 1|y) &= \int_0^1 P(\tilde{y} = 1|\theta, y)p(\theta|y)d\theta = \int_0^1 \theta p(\theta|y)d\theta \\ &= E(\theta|y); \text{(posterior expectation or mean)} \end{aligned}$$

- $\theta \sim U(0,1)$, $\theta|y \sim Beta(y+1, n+1-y)$, then $E(\theta|y) = \frac{y+1}{n+2}$.
- A more general prior is indeed, $\theta \sim Beta(a,b)$ or

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}; a>0, b>0$$

- The *Posterior distribution* for this prior can be obtained as,

$$p(\theta|y) \propto f(y|\theta)p(\theta) \propto \binom{n}{y}\theta^{y}(1-\theta)^{n-y}\theta^{a-1}(1-\theta)^{b-1}$$

- which implies

$$p(\theta|y) \propto \theta^{a+y-1}(1-\theta)^{b+n-y-1}.$$

- This is a Beta distribution, $Beta(a+y, n-y+b)$.

- **Conjugate prior:** Posterior distribution follows the same parametric form as the prior (also known as ”Data-augmentation prior”).
- The *prior-posterior* update is fairly simple.
- In the Binomial-Beta example, $a \to a + y$; $b \to b + n - y$.
- So a may be thought as the number of *successes* represented in the prior (b the number of failures).
- In this case,
$$E(\theta|y) = \frac{a+y}{a+b+n}.$$
- As $n \to \infty$, prior does not influence the posterior mean.
- The case $a = b = 1$ can be thought as a *Reference* case.

Illustrations

Example 1

- y is the number of girls in n recorded births. θ probability of a female birth.
- Assume that *births* are independent.
- Hypothesis $\theta < 0.5$. In Europe and accepted value is 0.485.
- If $\theta \sim U(0,1)$, then $p(\theta|y) = Beta(\theta|y+1, n-y+1)$.
- Posteriors for: a) $n = 5; y = 3$, b) $n = 20; y = 12$, c) $n = 100; y = 60$, d) $n = 1000; y = 600$.
- $p(\theta < 0.5|n = 5, y = 3)$?

Example 2

- Laplace collected data in Paris from 1745-1770.
- Result: 241,945 girls; 251,527 boys.
- $\theta \sim U(0, 1)$.
- $P(\theta < 0.5 | y = 241,945, n = 493,472)$?
- Posterior summaries based on the mean and SD of a Beta(a,b) distribution.
- If m denotes the mean, $m = \frac{a}{(a+b)}$; SD $= \sqrt{\frac{m(1-m)}{(a+b+1)}}$.
- We also consider a 95% probability interval.

Example 3

- *Placenta previa:* unusual condition in pregnancy. ”Placenta implanted low in the uterus”.
- Study in Germany: $n = 980$ births, $y = 437$ female.
- Is there evidence to think that $\theta < 0.485$?
- Use of different Beta priors based on 2 values,

$$m = \frac{a}{(a+b)}; s = (a+b)$$

  for m and s specified apriori.
- This is equivalent to,

$$a = ms; b = s(1-m)$$

- For a posterior $Beta(a+y, b+n-y)$, the sum of parameters is $a+b+n$ so $a+b$ can be interpreted as the equivalent sample size in the prior.

- How do we estimate the ratio of probabilities of male to female births?
- So the parameter of interest is $\phi = (1 - \theta)/\theta$.
- Difficult to find the posterior distribution of $p(\phi|y)$.
- An approximate solution with *Monte Carlo* simulation:
    - Generate 1000 (for example) draws from $\theta \sim Beta(a + y, b + n - y)$.
    - $\theta_1, \theta_2, \ldots, \theta_{1000}$.
    - With draws, compute values $\phi_i = (1 - \theta_i)/\theta_i; i = 1, 2, \ldots, 1000$.
    - Summarize values of $\phi_i, i = 1, 2, \ldots, 1000$ (histograms, empirical quantiles).