VE414 Lecture 3

Jing Liu

UM-SJTU Joint Institute

September 12, 2019

- However, Bayes did not focus on point estimates because he argued a point estimate fail to fully incorporate and reflect what we can learn from the data.

Q: What would you have as $\hat{p}_k$ and $\tilde{p}_k$ if you had the following data

$$X_2 = 2; \qquad X_{20} = 20; \qquad x_{200} = 200$$

Q: What should you report instead of point estimates?

Definition

In Bayesian analysis, a credible interval is an interval within which an unobserved parameter/variable value falls with a particular probability. Let $F_{Y|X}$ be the CDF,

$$F_{Y|X}(y_u \mid x) > F_{Y|X}(y_l \mid x)$$

then the interval $\mathcal{I} = [y_l, y_u]$ is a credible interval with coverage probability

$$F_{Y|X}(y_u \mid X) - F_{Y|X}(y_l \mid X)$$

If exactly $(\alpha/2)100\%$ of the posterior probability lies above and below $\mathcal{I}$, then $\mathcal{I}$ is known as the central credible interval with coverage probability of $1 - \alpha$.

```r
> # Prior is taken to be beta(1,1), i.e. uniform
> # Prior Hyperparameters
> a.prior = 1; b.prior = 1

> # Inverse of cdf, aka quantile function
> qbeta(0.5, a.prior, b.prior)
```

```
[1] 0.5
```

```r
> # Lower and upper limits
> l = qbeta(0.025, a.prior, b.prior)
> u = qbeta(0.975, a.prior, b.prior)

> # Print out the prior central credible interval
> cat(paste("(",l, sep = ""),
+     paste(u,")", sep =""), sep = ",")
```

```
(0.025,0.975)
```

```r
> # Extreme case
> k.vec = c(2, 20, 200)   # Number of trials
> x.vec = c(2, 20, 200)   # Data
> n = length(k.vec)       # Number of cases

> for (i in 1:n){
+   # Posterior hyperparameters
+   a.posterior = a.prior + x.vec[i]
+   b.posterior = b.prior + (k.vec[i]-x.vec[i])
+
+   l = qbeta(0.025, a.posterior, b.posterior)
+   u = qbeta(0.975, a.posterior, b.posterior)
+   cat(paste("(",l, sep = ""),
+       paste(u,")", sep =""),
+       "\n", sep = ",")  # New line
+ }
```

```
(0.292401773821287,0.991596241340387)
(0.83890238478092,0.998795116551636)
(0.981814749945614,0.99987404868883)
```

```
> # Simulation study of posterior CCI
> set.seed(414)
> n = 5000
> k.vec = 1:n
> true.prob = 0.65

> # n Bernoulli trials
> x.vec = rbinom(n, size = 1, prob = true.prob)
> head(x.vec)
```

```
[1] 0 1 0 1 1 0
```

```
> # forming simulated the data
> x.vec = cumsum(x.vec)
> head(x.vec)
```

```
[1] 0 1 1 2 3 3
```

- Notice how the posterior central credible interval shrinks as k grows.

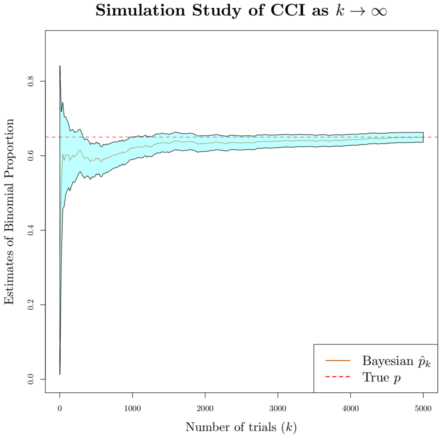
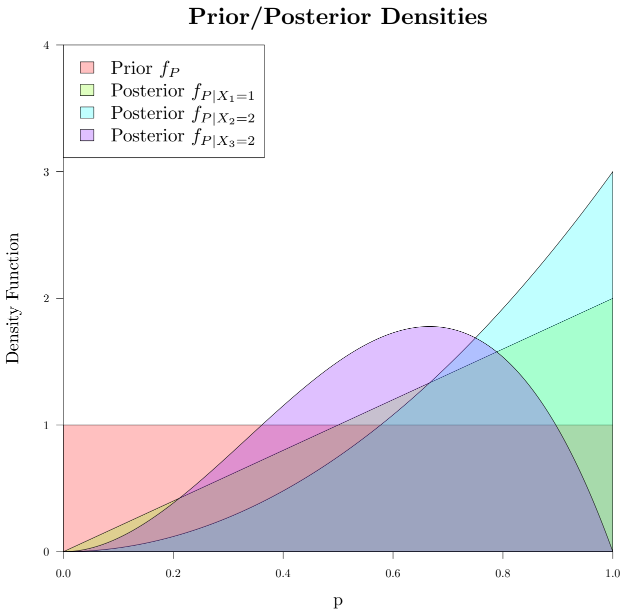


Figure: R Code: `central_credible_interval_binomial_beta.R`

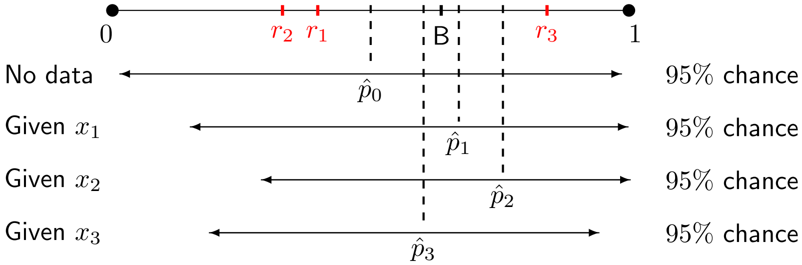
```
> # Back to Bayes' data
> k.vec = 1:3
> x.vec = c(1, 2, 2)
> n = length(k.vec)          # Number of cases

> for (i in 1:n){
+   # Posterior hyperparameters
+   a.posterior = a.prior + x.vec[i]
+   b.posterior = b.prior + (k.vec[i]-x.vec[i])
+   l = qbeta(0.025, a.posterior, b.posterior)
+   u = qbeta(0.975, a.posterior, b.posterior)
+   cat(paste("(",l, sep = ""),
+       paste(u,")", sep =""),
+       "\n", sep = ",")  # New line
+ }
```

```
(0.158113883008419,0.987420882906575)
(0.292401773821287,0.991596241340387)
(0.194120449683243,0.932414013511457)
```


Prior/Posterior Densities
Prior f_P
Posterior $f_{P|X_1=1}$
Posterior $f_{P|X_2=2}$
Posterior $f_{P|X_3=2}$
Density Function
p
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0

- In the Bayes' study, 95% credible intervals have a 95% chance of correctly capturing the true p, thus the black ball, at various stage of the experiment.



- Notice those intervals are different from typical confidence intervals (CI).

Q: What is the difference? How can we compute a typical confidence interval?

- Recall confidence intervals for binomial proportion p rely on the central limit theorem, and it is unreliable for a small number of trials. Hence suppose the experiment continues, and we have 3000 trials instead of just 3 trials.

- According to the central limit theorem, the following random variable

$$Z = (\tilde{p}_k - p) \bigg/ \sqrt{\frac{\tilde{p}_k (1 - \tilde{p}_k)}{k}}$$

follows a standard normal distribution as $k \to \infty$.

- Thus the following can be used as the $(1-\alpha)100\%$ confidence interval of p.

$$\left(\tilde{p}_k - z_{\alpha/2} \sqrt{\frac{\tilde{p}_k (1 - \tilde{p}_k)}{k}}, \tilde{p}_k + z_{\alpha/2} \sqrt{\frac{\tilde{p}_k (1 - \tilde{p}_k)}{k}} \right)$$

where $\alpha \in (0,1)$ and $z_{\alpha/2}$ is a real number such that $\Pr\left(Z \le -z_{\alpha/2}\right) = \frac{\alpha}{2}$.

- To understand a 95% confidence interval,
    1. consider $2^{16} = 65536$ samples of size 3000
    2. compute the 95% confidence interval for each sample
    3. check whether the true p is inside each of the confidence intervals

- Roughly 95% of the 95% confidence intervals contain the true p for large n

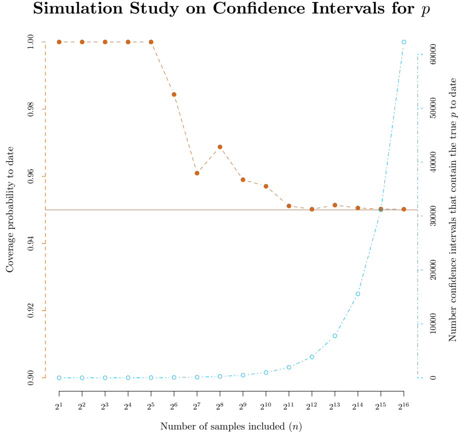


Figure: R Code: `bayes_ball_table_414.R`

- We can conduct something similar to a hypothesis test,

$$H_0 : p > 0.8$$

```
> a.prior = 1; b.prior = 1
> pbeta(0.8, a.prior, b.prior)      # cdf
```

```
[1] 0.8
```

```
> k.vec = 1:3; x.vec = c(1, 2, 2); n = length(k.vec)
> for (i in 1:n){
+   # Posterior hyperparameters
+   a.posterior = a.prior + x.vec[i]
+   b.posterior = b.prior + (k.vec[i]-x.vec[i])
+   cat(pbeta(0.8, a.posterior, b.posterior),
+       ";\t", sep = "")
+ }
```

```
0.64;   0.512;  0.8192;
```

Q: Is there any difference between Bayesian and Frequentist hypothesis testing?

- Recall p is considered to be a nonrandom parameter in the traditional sense, having an estimate of which allows us to predict X_k, for example, using MLE

$$\tilde{p}_3 = \frac{2}{3}$$

the following is sometime used to make point predictions

$$X_k^* \mid X_3 = 2 \sim \text{Binomial}\left(k, \frac{2}{3}\right)$$

since the exact sampling distribution of $\tilde{p}$ is not available as usual.

Q: Do you see the Bayesian approach offers an easy and nature way around?

$$\begin{aligned} f_{X_k^* \mid X_3=2}\left(x_k^* \mid 2\right) &= \int_{-\infty}^{\infty} f_{\{X_k^*, P\} \mid X_3=2}\left(x_k^*, p \mid 2\right) dp \\ &= \int_{-\infty}^{\infty} f_{X_k^* \mid P}\left(x_k^* \mid p\right) f_{P \mid X_3=2}\left(p \mid 2\right) dp \end{aligned}$$

Definition

In Bayesian analysis, given a prior of some parameter Y

$$f_Y(y)$$

and a likelihood of some observed data $X = x$,

$$\mathcal{L}(y; x) = f_{X|Y}(x \mid y)$$

the posterior predictive distribution of a future observation, X^*, is given by

$$\begin{aligned} f_{X^*|X}(x^* \mid x) &= \int_{-\infty}^{\infty} f_{\{X^*,Y\}|X}(x^*, y \mid x)\, dy \\ &= \int_{-\infty}^{\infty} f_{X^*|\{Y,X\}}(x^* \mid y, x)\, f_{Y|X}(y \mid x)\, dy \\ &= \int_{-\infty}^{\infty} f_{X^*|Y}(x^* \mid y)\, f_{Y|X}(y \mid x)\, dy \end{aligned}$$

Similarly, the prior predictive distribution $f_{X^*}(x^*) = \int_{-\infty}^{\infty} f_{X^*|Y}(x^* \mid y)\, f_Y(y)\, dy$

Q: In terms of Bayes' original problem, given $X_3 = 2$, the binomial likelihood and the uniform prior, what is the posterior predictive distribution for X_3^*?

$$
\begin{aligned}
f_{X_k^*|X_3=2} &= \int_0^1 f_{X_k^*|P} \cdot f_{P|X_3=2}\, dp \\
&= \int_0^1 \frac{k^*!}{x_k^*!(k^*-x_k^*)!} p^{x_k^*}(1-p)^{k^*-x_k^*} \qquad \text{where } k^* = 3 \\
&\qquad \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1}(1-p)^{\beta^*-1}\, dp \quad \text{and } \alpha^* = 3;\ \beta^* = 2 \\
&= \frac{k^*!}{x_k^*!(k^*-x_k^*)!} \cdot \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \int_0^1 p^{x_k^*+\alpha^*-1}(1-p)^{k^*-x_k^*+\beta^*-1}\, dp \\
&= \frac{k^*!}{x_k^*!(k^*-x_k^*)!} \cdot \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \cdot \frac{\Gamma(x_k^*+\alpha^*)\Gamma(k^*-x_k^*+\beta^*)}{\Gamma(k^*+\alpha^*+\beta^*)}
\end{aligned}
$$

- This distribution of X_k^* is known as the Beta-Binomial distribution.