1 Constructing the posterior distribution: Binomial Case

Construct the posterior distribution by combining the binomial likelihood with a beta prior.

- Likelihood of *actual* experiment: $p(y|\theta) = L(\theta|y) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$ (binomial)
- Prior distribution: $p(\theta) = \frac{1}{B(\alpha_0,\beta_0)}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$ (beta distribution)
- Bayes' theorem: $p(\theta|y) = \frac{L(\theta|y)p(\theta)}{f(y)} = \frac{L(\theta|y)p(\theta)}{\int L(\theta|y)p(\theta)d\theta}$
  - Numerator:

$$\begin{aligned} L(\theta|y)p(\theta) &= \binom{n}{y}\theta^y(1-\theta)^{n-y} \times \frac{1}{B(\alpha_0,\beta_0)}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1} \\ &= \binom{n}{y}\frac{1}{B(\alpha_0,\beta_0)}\theta^{y+\alpha_0-1}(1-\theta)^{n-y+\beta_0-1} \end{aligned}$$

  - Denominator:

$$\begin{aligned} f(y) &= \int L(\theta|y)p(\theta)d\theta \\ &= \binom{n}{y}\frac{1}{B(\alpha_0,\beta_0)}\int_0^1 \theta^{\alpha_0+y-1}(1-\theta)^{\beta_0+n-y-1}d\theta \\ &= \binom{n}{y}\frac{B(\alpha_0+y,\beta_0+n-y)}{B(\alpha_0,\beta_0)} \end{aligned}$$

  The last step is based on the fact that the area under the curve (AUC) of the density of a beta-distribution must be equal to 1 (since it is a density). Indeed, the AUC of a beta(α,β)-density is

$$\int \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = 1,$$

  and thus

$$\int \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = B(\alpha,\beta).$$

If you replace $\alpha \equiv \alpha_0 + y$ and $\beta \equiv \beta_0 + n - y$, you see that the integral can be replaced by the beta-function.

– Thus, the posterior distribution is (numerator divided by denominator):

$$p(\theta|y) = \frac{1}{B(\alpha_0 + y, \beta_0 + n - y)} \theta^{\alpha_0 + y - 1} (1 - \theta)^{\beta_0 + n - y - 1}$$

corresponding to a beta-distribution with parameters $\alpha_0 + y$ and $\beta_0 + n - y$.

2 Plot binomial likelihood, beta prior and beta posterior in R

In R, you can first specify the number of successes, number of trials and the parameters of the beta-prior by the following code. For example:

```
x<-3         # number of successes
N<-20        # number of trials
alpha<-10    # alpha-parameter of beta-prior
beta<-20     # beta-parameter of beta-prior
```

Since likelihood, prior and posterior are known distributions, we specify the binomial likelihood (`likl`), the prior distribution (`prior`) and the posterior distribution (`posterior`) using the following R-functions:

```
theta<-seq(0,1,0.01)
# prior distribution
prior<-dbeta(x=theta,shape1=alpha,shape2=beta)
# scaled likelihood function
likl<-dbeta(x=theta,shape1=x+1,shape2=N-x+1)
# posterior distribution
posterior<-dbeta(theta,alpha+x,beta+N-x)
```

Note that in the above code we calculate the scaled likelihood value $L(\theta|x)/\int L(\theta|x)d\theta$, the prior function $p(\theta)$ and posterior function $p(\theta|x)$ in different values of θ. From the course notes, we know that the beta distribution $Beta(x+1, N-x+1)$ is proportional to the binomial likelihood with x successes in $(x+1)+(N-x+1)-2=N$ operations (since the beta distribution is obtained by standardizing the binomial likelihood). In order to make a plot of prior, scaled likelihood and posterior as function of the parameter θ, we use the `plot` and `lines` functions in R:

```
par(lwd=2)
plot(theta,prior,type="l",lty=1,ylim=c(0,8))
lines(theta,likl,lty=2,col="blue")
lines(theta,posterior,lty=4,col="red")
```

A legend can be added to the plot by the following code:

```
legend(0.6,8,c("prior","scaled likelihood", "posterior"),
       lty=c(1,2,3),col=c("black","blue","red"))
```

3 Calculations Posterior Distribution: Normal Case

- Likelihood of *actual* experiment: $L(\mu|y) \propto \exp\left(-\frac{1}{2}\left(\frac{\mu-\bar{y}}{\sigma/\sqrt{n}}\right)^2\right)$ (normal)

- Prior distribution: $p(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right)$ (normal distribution)

- Bayes' theorem: $p(\mu|y) = \frac{L(\mu|y)p(\mu)}{f(y)} = \frac{L(\mu|y)p(\mu)}{\int L(\mu|y)p(\mu)d\mu}$

  – Posterior:

$$
\begin{aligned}
&\propto \quad L(\mu|y)p(\mu) \\
&\propto \quad \exp\left(-\frac{1}{2}\left[\left(\frac{\mu-\bar{y}}{\sigma/\sqrt{n}}\right)^2 + \left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right]\right) \\
&\propto^* \quad \exp\left(-\frac{1}{2}\left[\left(\frac{\mu^2-2\mu\bar{y}}{\sigma^2/n}\right) + \left(\frac{\mu^2-2\mu\mu_0}{\sigma_0^2}\right)\right]\right) \\
&= \quad \exp\left(-\frac{1}{2}\left[\left(\frac{1}{\sigma^2/n}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\mu\left(\frac{\bar{y}}{\sigma^2/n}+\frac{\mu_0}{\sigma_0^2}\right)\right]\right)
\end{aligned}
$$

  * Note that in this step we are working out the squares, but drop the factors that do not contain the parameter μ, since it is sufficient to calculate the product of likelihood with prior up to a constant (proportionality factor).

- We can recognise that this product is, up to a proportionality contant, equal to a normal density. Indeed, a normal density $x \sim N(\mu, \sigma)$ has density

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \propto \exp\left[-\frac{1}{2\sigma^2}\left(x^2-2x\mu\right)^2\right]$$

From the previous expression we now find that

$$L(\mu|y)p(\mu) \propto \exp\left(-\frac{1}{2\bar{\sigma}^2}\left[\mu^2 - 2\mu\bar{\mu}\right]\right)$$

with

$$\bar{\sigma}^2 = \left(\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2} \right)^{-1}$$

and

$$\bar{\mu} = \left(\frac{\bar{y}}{\sigma^2/n} + \frac{\mu_0}{\sigma_0^2} \right) / \bar{\sigma}^2$$

- As a result, we conclude that

$$p(\mu|y) \equiv N(\bar{\mu}, \bar{\sigma}^2)$$

4 Alternative interpretation

$$\bar{\mu} = \frac{w_0}{w_0 + w_1}\mu_0 + \frac{w_1}{w_0 + w_1}\bar{y}$$

with $w_0 = \frac{n_0}{\sigma^2}$ and $w_1 = \frac{n}{\sigma^2}$.

Thus,

$$\begin{aligned}
\bar{\mu} &= \frac{w_0}{w_0 + w_1}\mu_0 + \frac{w_1}{w_0 + w_1}\bar{y} \\
&= \frac{\frac{n_0}{\sigma^2}}{\frac{n_0}{\sigma^2} + \frac{n}{\sigma^2}}\mu_0 + \frac{\frac{n}{\sigma^2}}{\frac{n_0}{\sigma^2} + \frac{n}{\sigma^2}}\bar{y} \\
&= \frac{n_0}{n_0 + n}\mu_0 + \frac{n}{n_0 + n}\bar{y}
\end{aligned}$$

And also,

$$\bar{\sigma}^2 = \left(\frac{n_0}{\sigma^2} + \frac{n}{\sigma^2}\right)^{-1} = \left(\frac{n_0 + n_1}{\sigma^2}\right)^{-1} = \frac{\sigma^2}{n_0 + n_1}$$