DDA 4010 Assignment 1 赵松霖 120090346

1.

$E(a_1Y_1 + a_2Y_2) = a_1E(Y_1) + a_2E(Y_2) = a_1u_1 + a_2u_2$

$Var(a_1Y_1 + a_2Y_2) = E([a_1Y_1 + a_2Y_2 - E(a_1Y_1 + a_2Y_2)]^2)$

$= E[(a_1Y_1 - a_1u_1 + a_2Y_2 - a_2u_2)^2]$

$= E[(a_1Y_1 - a_1u_1)^2 + (a_2Y_2 - a_2u_2)^2 + 2a_1a_2(Y_1-u_1)(Y_2-u_2)]$

$= a_1^2E[(Y_1-u_1)^2] + a_2^2E[(Y_2-u_2)^2] + 2a_1a_2\,cov(Y_1,Y_2)$

$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2$

$E(a_1Y_1 - a_2Y_2) = a_1E(Y_1) - a_2E(Y_2) = a_1u_1 - a_2u_2$

$Var(a_1Y_1 - a_2Y_2) = E([a_1Y_1 - a_2Y_2 - E(a_1Y_1 - a_2Y_2)]^2)$

$= E[(a_1Y_1 - a_1u_1)^2 + (a_2Y_2 - a_2u_2)^2 - 2a_1a_2(Y_1-u_1)(Y_2-u_2)]$

$= a_1^2E[(Y_1-u_1)^2] + a_2^2E[(Y_2-u_2)^2] - 2a_1a_2\,cov(Y_1,Y_2)$

$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2$

2.

a) $P(X=0, Y=0) = P(Y=0|X=0)P(X=0) = 0.5 \cdot 0.4 = 0.2$

$P(X=0, Y=1) = P(Y=1|X=0)P(X=0) = 0.5 \cdot 0.6 = 0.3$

$P(X=1, Y=0) = P(Y=0|X=1)P(X=1) = 0.5 \cdot 0.6 = 0.3$

$P(X=1, Y=1) = P(Y=1|X=1)P(X=1) = 0.5 \cdot 0.4 = 0.2$

		Y	
		0	1
X	0	0.2	0.3
	1	0.3	0.2

b) Since $Y|X=0 \sim Bernoulli(0.6)$

$Y|X=1 \sim Bernoulli(0.4)$

$E(Y) = E(E(Y|X)) = E(Y|X=0)P(X=0) + E(Y|X=1)P(X=1)$

$= 0.6 \times 0.5 + 0.4 \times 0.5$

$= 0.5$

c) $Var(Y|X=0) = 0.6 \times (1-0.6) = 0.24$

$Var(Y|X=1) = 0.4(1-0.4) = 0.24$

Since $P(Y=0) = \sum_x P(X=x, Y=0) = 0.5$

$P(Y=1) = \sum_x P(X=x, Y=1) = 0.5$

$\therefore \; Var(Y) = E((Y-\mu_Y)^2)$

$= (0-0.5)^2 P(Y=0) + (1-0.5)^2 P(Y=1)$

$= 0.5 \times 0.25 + 0.5 \times 0.25$

$= 0.25$

Given $X=0$ or $X=1$, There is less uncertainty for Y, So $Var(Y|X=0)$ and $Var(Y|X=1)$ are less than $Var(Y)$

d) $P(X=0 \mid Y=1) = \dfrac{P(Y=1 \mid X=0)\, P(X=0)}{P(Y=1)}$

$= \dfrac{0.6 \times 0.5}{0.5}$

$= 0.6$

3.

We interpret $p(E)$ as: willing to give $P(E)$ to someone, provided they give me \$1 if E occurs.

a) Since E sometimes occurs, in order not to lose money, I'm willing to give someone $p(E)$, which is ≤ 1.
At the extreme case, E must occur, then $p(E) \leq 1$ can guarantee that I will not lose money.

b) E^c: willing to give $p(E^c)$ to someone, provided they give you \$1 if E does not occur.

Then, either E will occur or not occur.

$\therefore$ the other person will always given \$1 to me.

Since each time I am willing to give someone $p(E) + p(E^c)$ each time,

then $p(E) + p(E^c) \leq 1$

By another definition: $p(E) + p(E^c) \geq 1$ by same logic.

$\therefore \; p(E) + p(E^c) = 1$

4.

a) $Y_i \mid \theta \sim$ iid Bernoulli (θ)

$Pr(Y_1 = y_1, \ldots, Y_{100} = y_{100} \mid \theta) = P(Y_1 = y_1 \mid \theta)\, P(Y_2 = y_2 \mid \theta) \cdots P(Y_{100} = y_{100} \mid \theta)$

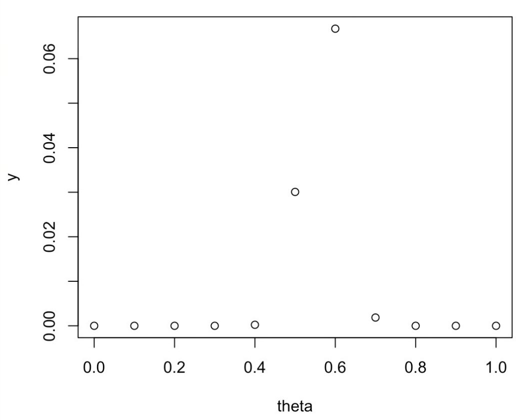
$= \theta^{y_1}(1-\theta)^{1-y_1} \cdots \theta^{y_{100}}(1-\theta)^{1-y_{100}}$

$= \theta^{\Sigma y_i}(1-\theta)^{100-\Sigma y_i}$

$$Pr(\Sigma Y_i = y \mid \theta) = \binom{100}{y} \theta^{y} (1-\theta)^{100-y}$$

b) $p(\Sigma Y_i = 57 \mid \theta) = \binom{100}{57} \theta^{57} (1-\theta)^{100-57}$

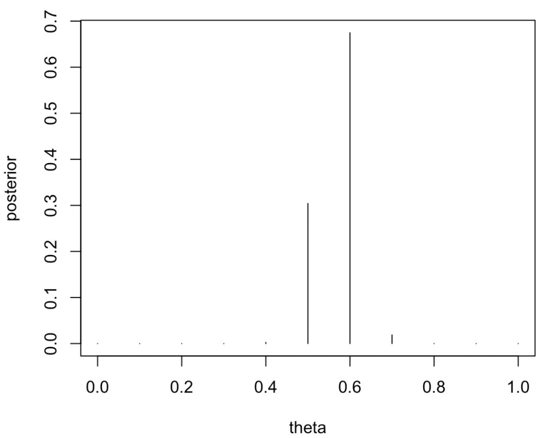
```
> theta = seq(0,1,by=0.1)
> y = choose(100, 57) * theta^57 * (1-theta)^43
> plot(theta, y)
```



c) $p(\theta \mid \Sigma Y_i = 57) \propto p(\Sigma Y_i = 57 \mid \theta)\, p(\theta)$

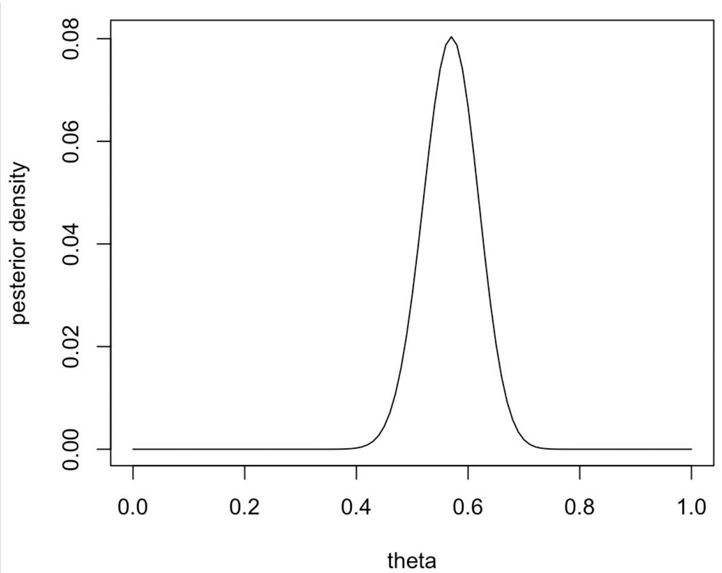
$$\propto \theta^{57} (1-\theta)^{43}$$

```
> theta = seq(0,1,by=0.1)
> prob = theta^57 * (1-theta)^43
> posterior = prob / sum(prob)
> plot(theta, posterior, type='h')
```

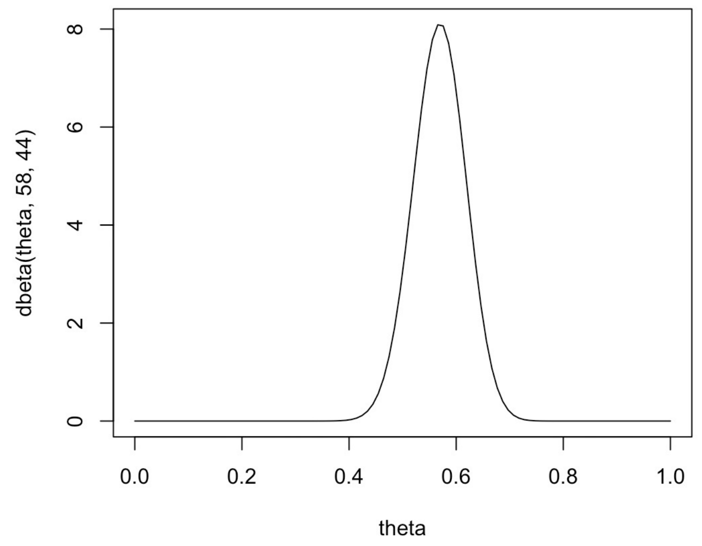


d) $p(\theta) \cdot Pr(\Sigma Y_i = 57 \mid \theta) = \binom{100}{57} \theta^{57} (1-\theta)^{100-57}$

```
> eq = function(theta){choose(100, 57) * theta^57 * (1-theta)^43}
> curve(eq, from=0, to=1, xlab="theta", ylab="pesterior density")
```



e)
```
> theta = seq(0,1, length=100)
> plot(theta, dbeta(theta, 58, 44), type='l')
```



Shapes of the four plots are same, but with different scale.

(b) is a discrete plot of $p(\text{data}|\theta)$

(c) is a discrete plot of the posterior $p(\theta|\text{data})$

(d) is a posterior density with prior $U(0,1)$

(e) is the posterior with prior $U(0,1)$

5. $a = \theta_0 n_0$

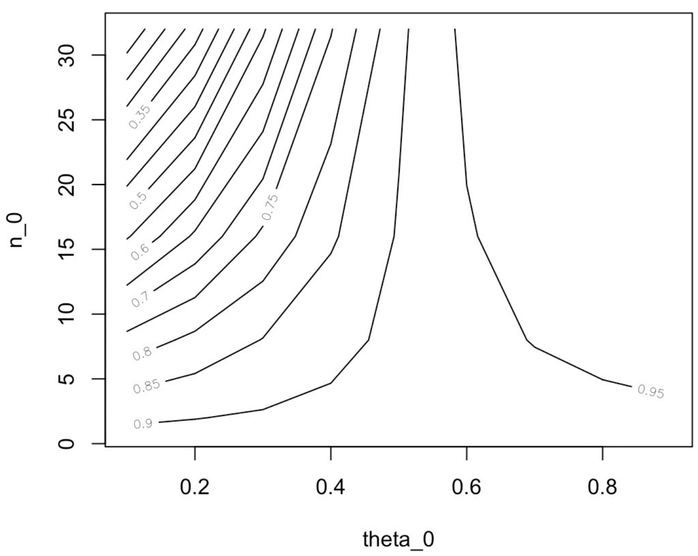
$b = (1-\theta_0) n_0$

$$p(\theta|\Sigma Y_i = 57) = \frac{p(\Sigma Y_i = 57|\theta)\, p(\theta)}{p(\Sigma Y_i = 57)}$$

$$\propto \theta^{57}(1-\theta)^{43}\, \theta^{a-1}(1-\theta)^{b-1}$$

$$= \theta^{57+a-1}(1-\theta)^{43+b-1}$$

$$\therefore p(\theta|\Sigma Y_i = 57) = \frac{\Gamma(100+a+b)}{\Gamma(57+a)\Gamma(43+b)}\theta^{57+a-1}(1-\theta)^{43+b-1} \sim \text{beta}(57+a, 43+b)$$




```
theta_0 <- seq(0.1,0.9,by=0.1)
n_0 <- c(1,2,8,16,32)
a = outer(theta_0,n_0,FUN="*")
b = outer((1-theta_0),n_0,FUN="*")
probability = 1 - pbeta(0.5,57+a,43+b)
contour(theta_0, n_0, probability, nlevels = 20, xlab="theta_0", ylab="n_0")
```

θ_0 can be interpreted as prior expectation

n_0 can be interpreted as prior sample size

as θ_0 increases, $p(\theta > 0.5|\Sigma Y_i = 57)$ increases.

as $n_0 \downarrow$, the outcome $\Sigma y_i = 57$ tends to dominate the result. as $n_0 \uparrow$, the prior dominates the result.

When n is small, $p(\theta > 0.5|\Sigma y_i = 57)$ is large, so we tend to believe that $\theta > 0.5$

6.

a) $p(\theta) = 1 = Beta(1,1)$

$p(y_1|\theta) = \binom{15}{y_1} \theta^{y_1} (1-\theta)^{15-y_1}$

$p(\theta|y_1) \propto p(\theta)\, p(y_1|\theta)$

$\propto \theta^{y_1}(1-\theta)^{15-y_1} = \theta^2(1-\theta)^{13}$

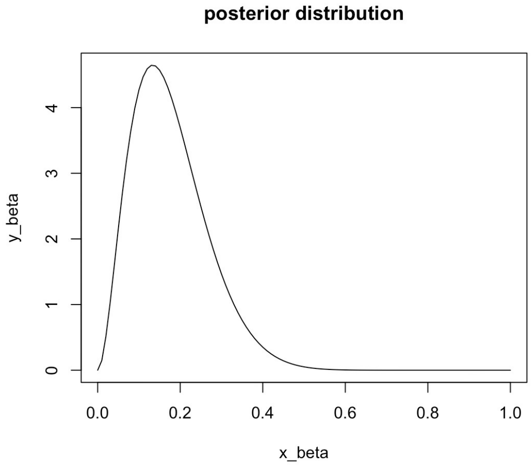
$\therefore p(\theta|y_1) \sim beta(y_1+1, 16-y_1) = beta(3,14)$

$E(\theta|y_1) = \frac{3}{17}$

$Mode(\theta|y_1) = \frac{2}{15}$

$Var(\theta|y_1) = \frac{3\times14}{18\times17^2} = 0.09$

```
x_beta <- seq(0, 1, by = 0.01 )
y_beta <- dbeta(x_beta, shape1 = 3, shape2 = 14)
plot(x_beta,y_beta,type="l")
```



b) i, $P(Y_2=y_2 \mid Y_1=2)$

$= \int_0^1 P(Y_2=y_2, \theta \mid Y_1=2)\, d\theta$

$= \int_0^1 P(Y_2=y_2 \mid \theta, Y_1=2)\, P(\theta|Y_1=2)\, d\theta$

$\therefore P(Y_2=y_2 \mid \theta, Y_1=2) = P(Y_2=y_2|\theta)$

$\frac{P(Y_2=y_2, Y_1=2 \mid \theta)}{P(Y_1=2|\theta)} = P(Y_2=y_2|\theta)$

$\therefore P(Y_2=y_2, Y_1=2|\theta) = P(Y_2=y_2|\theta)\, P(Y_1=2|\theta)$

$\therefore$ $Y_2=y_2$ and $Y_1=2$ are conditional independent given θ.

ii. $Pr(Y_2=y_2 \mid Y_1=2)$

$= \int_0^1 Pr(Y_2=y_2|\theta)\, p(\theta|Y_1=2)\, d\theta$

$= \int_0^1 \binom{278}{y_2} \theta^{y_2}(1-\theta)^{278-y_2} \frac{16!}{2!13!} \theta^2 (1-\theta)^{13}\, d\theta$

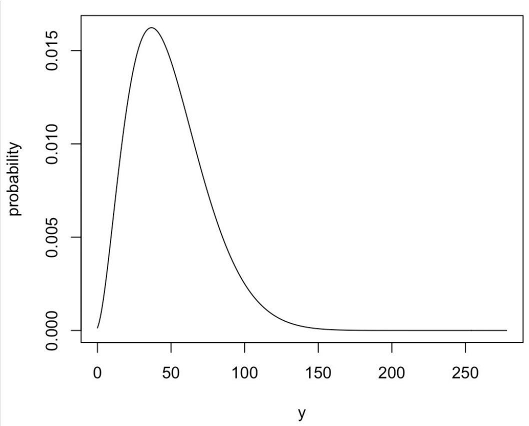
iii. $Pr(Y_2 = y_2 \mid Y_1 = 2)$

$$= \binom{278}{y_2}\left(\frac{16!}{2!13!}\right)\int_0^1 \theta^{y_2+2}(1-\theta)^{291-y_2}\,d\theta$$

$$= \binom{278}{y_2}\frac{\Gamma(17)}{\Gamma(3)\Gamma(14)}\frac{\Gamma(y_2+3)\Gamma(292-y_2)}{\Gamma(295)}$$

c)

```
y = seq(0,278,by=1)
probability = choose(278, y)*16*choose(15, 2)/choose(294,y+2)/(292-y)
plot(y, probability, type="l")
```



$E(Y_2 \mid Y_1 = 2) = 49.05882$

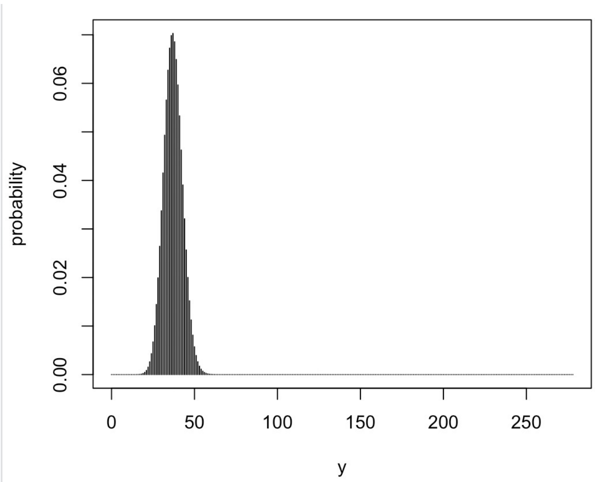
$Sd(Y_2 \mid Y_1 = 2) = 25.73196$

```
> weighted.mean(y, probability)
[1] 49.05882
> (weighted.mean((y-49.05882)^2, probability))^0.5
[1] 25.73196
```

d) $Pr(Y_2 = y_2 \mid \theta = \hat{\theta}) = \binom{278}{y_2}\hat{\theta}^{y_2}(1-\hat{\theta})^{278-y_2}$

$$= \binom{278}{y_2}\left(\frac{2}{15}\right)^{y_2}\left(\frac{13}{15}\right)^{278-y_2}$$

```
y = seq(0,278,by=1)
probability = dbinom(y, 278, 2/15)
plot(y, probability, type="h")
```



$E(Y_2 \mid \theta = \hat{\theta}) = 278 \cdot \frac{2}{15} = 37.0667$

$Sd(Y_2 \mid \theta = \hat{\theta}) = \sqrt{278 \cdot \frac{2}{15} \cdot \frac{13}{15}} = 5.667843$

(d) is the froquentists' view while (c) is bayesians' view.

(d) has smaller standard deviation while (c) has larger standard deviation

I would use (c) because (d) fixes the value of $\theta = \hat{\theta}$, which is the MLE or mode.

But it's better not to fix $\theta = \hat{\theta}$, instead consider all θ using $P(Y_2 | Y_1) = \int_0^1 P(Y_2, \theta | Y_1)$.