Problem Set 3 -- Solutions

Q1.

a.

$$\theta_1 = 0.25, p(\theta_1) = 0.25, \theta_2 = 0.5, p(\theta_2) = 0.5, \theta_3 = 0.75, p(\theta_3) = 0.25$$

$$p(H = 50, T = 50) = \sum_{i=1}^{3} p(H = 50, T = 50|\theta_i)p(\theta_i)$$

$$= p(H = 50, T = 50|0.25)p(0.25) + p(H = 50, T = 50|0.5)p(0.5) + p(H = 50, T = 50|0.75)p(0.75)$$

$$= 0.25\ Binomial(H = 50, T = 50|0.25) + 0.5\ Binomial(H = 50, T = 50|0.5) + 0.25\ Binomial(H = 50, T = 50|0.75)$$

$$\approx 0.039795$$

I used the following Octave code to compute the solution:

```
0.25 * binopdf(50,100,0.25) + 0.5 * binopdf(50,100,0.5) + 0.25 * binopdf(50,100,0.75)
```

b.

$$p(\theta = 0.5|H = 50, T = 50) = \frac{p(H = 50, T = 50|\theta = 0.5)p(\theta = 0.5)}{p(H = 50, T = 50)}$$

$$= \frac{0.5\ Binomial(H = 50, T = 50|0.5)}{p(H = 50, T = 50)}$$

$$= \frac{0.5\ Binomial(H = 50, T = 50|0.5)}{0.039795}$$

$$\approx \frac{0.5 \times 0.079589}{0.039795} \approx 1$$

In Octave: `0.5 * binopdf(50,100,0.5)/0.039795`

Q2.

a. The solution can be computed using the Beta-Binomial distribution.

$$BetaBinomial(H, T|\alpha, \beta) = \binom{H+T}{H} \frac{B(\alpha + H, \beta + T)}{B(\alpha, \beta)}$$

$$= \binom{100}{50} \frac{B(5 + 50, 10 + 50)}{B(5,10)}$$

$$= 0.012711$$

Computed using Octave: `nchoosek(100,50)*beta(55,60)/beta(5,10)`

b. This question is potentially ambiguous.

If the value of the *cdf* (cumulative density function) is desired then the answer is 0. Since, *cdf* (i.e. area under the *pdf*) of a real-valued probability distribution, at a fixed location, is 0.

If the value of the *pdf* (probability density function) is desired then:

The posterior distribution for θ is given by:

$$
\begin{aligned}
p(\theta = 0.5|\alpha, \beta, H = 50, T = 50) &= \frac{p(H,T|\theta)p(\theta|\alpha,\beta)}{p(H,T|\alpha,\beta)} \\
&= \frac{Binomial(H,T|\theta)Beta(\theta|\alpha,\beta)}{BetaBinomial(H,T|\alpha,\beta)} \\
&= \frac{Binomial(H,T|\theta)Beta(\theta|\alpha,\beta)}{0.012711} \\
&= 7.6510
\end{aligned}
$$

c. Since the prior has $\boldsymbol{\alpha = 5}$ and $\boldsymbol{\beta = 10,}$ the Beta distribution after seeing 50 heads and 50 tails has parameters $\boldsymbol{\alpha = 5 + 50 = 55}$ and $\boldsymbol{\beta = 10 + 50 = 60}$
So what we want is:

$$\int_{0.4}^{0.6} Beta(\boldsymbol{\theta}|\boldsymbol{\alpha = 55, \beta = 60})d\theta$$

This is not an integral you can work out by hand! Fortunately octave has the cumulative density function for the Beta distribution built-in, so we can do:

```
octave> betacdf(0.6,55,60)-betacdf(0.4,55,60)
ans =  0.95043
```

Answer = 0.95043
Discussion:

Thus 95% of the probability mass is concentrated between 0.4 and 0.6 which is to be expected since the coin tosses were perfectly balanced **even after 100 throws** and the effect of the priors is having minimal effect in the light of strong evidence.

We can check the probability mass at other intervals by altering the values in the betacdf function (highlighted in blue):

0.0-0.3 : 0.000032
0.3-0.4 : 0.045287
0.4-0.6: 0.95044
0.6-0.8 : 0.004253

showing that the pdf is highly peaky and the probability mass quickly tails off as you move away from 0.5.

Q3. Compute the expected value i.e. E[θ] of the posterior of the binomial distribution under the beta prior. See the lecture slides for explanation of this distribution. If the coin was thrown $(c_1 + c_2)$ times giving c_1 heads and c_2 tails, explain in plain English, what does E[θ] mean?

The posterior distribution for the Binomial under beta prior is given by:

$$P\big(\theta\big|\alpha,\beta,c=(c_1,c_2)\big)\sim Beta(\alpha+c_1,\beta+c_2)$$
$$=\frac{\Gamma(c_1+c_2+\alpha+\beta)}{\Gamma(\alpha+c_1)\Gamma(\beta+c_2)}\theta^{\alpha+c_1-1}(1-\theta)^{\beta+c_2-1}$$

$$E[\theta]=\int\theta\,p\big(\theta\big|\alpha,\beta,c=(c_1,c_2)\big)\,d\theta$$
$$\int\theta\frac{\Gamma(c_1+c_2+\alpha+\beta)}{\Gamma(\alpha+c_1)\Gamma(\beta+c_2)}\theta^{\alpha+c_1-1}(1-\theta)^{\beta+c_2-1}d\theta$$
$$=\frac{\Gamma(c_1+c_2+\alpha+\beta)}{\Gamma(\alpha+c_1)\Gamma(\beta+c_2)}\int\theta^{\alpha+c_1+1-1}(1-\theta)^{\beta+c_2-1}d\theta$$
$$=\frac{\Gamma(c_1+c_2+\alpha+\beta)}{\Gamma(\alpha+c_1)\Gamma(\beta+c_2)}\frac{\Gamma(\alpha+c_1+1)\Gamma(\beta+c_2)}{\Gamma(c_1+c_2+\alpha+\beta+1)}$$
$$=\frac{\Gamma(c_1+c_2+\alpha+\beta)}{\Gamma(\alpha+c_1)}\cdot\frac{(\alpha+c_1)\Gamma(\alpha+c_1)}{(c_1+c_2+\alpha+\beta)\Gamma(c_1+c_2+\alpha+\beta)}$$
$$=\frac{\alpha+c_1}{c_1+c_2+\alpha+\beta}$$

since $\Gamma(\mathrm{N}+1)=\mathrm{N}\cdot\Gamma(\mathrm{N})$.

Meaning of E[θ]:

So if the coin shows c_1 heads from (c_1+c_2) flips then the averaged bias θ of the coin is given by:

$$\frac{\alpha+c_1}{c_1+c_2+\alpha+\beta}$$

Q4. Suppose that $\boldsymbol{X_1},\dots,\boldsymbol{X_n}$ are independent identically distributed samples from a univariate Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma^2}$.

a. Find the ML estimate of the mean $\boldsymbol{\mu}$
b. Find the MAP estimate of the mean $\boldsymbol{\mu}$ when the mean $\boldsymbol{\mu}$ is distributed according to a Gaussian prior with mean $\boldsymbol{\alpha}$ and variance $\boldsymbol{\sigma^2}$

Comment on the difference.

Solution

a.

$$\boldsymbol{\mu_{MLE}}=\underset{\mu}{\mathbf{argmax}}\,\boldsymbol{P(X_1=x_1,\dots,X_n=x_n|\mu,\sigma)}$$
$$=\underset{\mu}{\mathbf{argmax}}\left(\prod_{i=1}^{n}\frac{\mathbf{1}}{\sqrt{\mathbf{2}\pi\sigma}}\boldsymbol{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$$
$$=\underset{\mu}{\mathbf{argmax}}\left(\left(\frac{\mathbf{1}}{\mathbf{2}\pi\sigma^2}\right)^{n/2}\boldsymbol{e}^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right)$$
$$=\underset{\mu}{\mathbf{argmax}}\left(\mathbf{log}\left(\left(\frac{\mathbf{1}}{\mathbf{2}\pi\sigma^2}\right)^{n/2}\boldsymbol{e}^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right)\right)$$
$$=\underset{\mu}{\mathbf{argmax}}\left(\mathbf{log}\left(\frac{\mathbf{1}}{\mathbf{2}\pi\sigma^2}\right)^{n/2}+\mathbf{log}\,\boldsymbol{e}^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right)$$

$$= \underset{\mu}{\operatorname{argmax}} \left(\log \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} - \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \right)$$

Taking the derivative with respect to **μ**, and setting it equal to **0** we get:

$$0 = \frac{\partial}{\partial\mu} \left(\log \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} - \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \right)$$
$$0 = 0 - \frac{2\sum_{i=1}^{n} (x_i - \mu)\,(-1)}{2\sigma^2}$$
$$0 = \frac{\left(\sum_{i=1}^{n} x_i\right) - n\mu}{\sigma}$$
$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

b. $$\mu_{MAP} = \underset{\mu}{\operatorname{argmax}}\, P(X_1 = x_1, \dots, X_n = x_n | \mu, \sigma) P(\mu)$$
$$= \underset{\mu}{\operatorname{argmax}} \left(\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\alpha-\mu)^2}{2\sigma^2}} \right)$$
$$= \underset{\mu}{\operatorname{argmax}} \left(\left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n+1}{2}} e^{-\frac{(\alpha-\mu)^2 + \sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \right)$$
$$= \underset{\mu}{\operatorname{argmax}} \left(\log \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n+1}{2}} + \log e^{-\frac{(\alpha-\mu)^2 + \sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \right)$$
$$= \underset{\mu}{\operatorname{argmax}} \left(\log \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n+1}{2}} - \frac{(\alpha - \mu)^2 + \sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} \right)$$

Taking the derivative with respect to μ equal to 0:

$$0 = \frac{\partial}{\partial\mu} \left(\log \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n+1}{2}} - \frac{(\alpha - \mu)^2 + \sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} \right)$$
$$0 = 0 - \frac{-2(\alpha - \mu) - 2\sum_{i=1}^{n}(x_i - \mu)}{2\sigma^2}$$
$$0 = \alpha - \mu + \left(\sum_{i=1}^{n} x_i \right) - n\mu$$
$$\mu = \frac{\alpha + \left(\sum_{i=1}^{n} x_i\right)}{n+1}$$

Using a prior with the same **σ** is equivalent to having an additional sample with value equal to the prior's mean.

Q5. Prove that if there are $\boldsymbol{n}$ successes from $\boldsymbol{N}$ Bernoulli trials, then the maximum likelihood estimator (**MLE**) of the probability of success $\boldsymbol{\theta = n/N}$.

Solution

$$
\begin{aligned}
\boldsymbol{\theta_{MLE}} &= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\, \boldsymbol{p(x|\theta)} \\
&= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\ \boldsymbol{\theta^{n}(1-\theta)^{N-n}} \\
&= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\ \boldsymbol{log(\theta^{n}(1-\theta)^{N-n})} \\
&= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\ (\boldsymbol{n}\,\mathbf{log}\,\boldsymbol{\theta} + (\boldsymbol{N-n})\,\mathbf{log}(\mathbf{1}-\boldsymbol{\theta}))
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{0} &= \frac{\boldsymbol{\partial}}{\boldsymbol{\partial\theta}}(\boldsymbol{n}\,\mathbf{log}\,\boldsymbol{\theta} + (\boldsymbol{N-n})\,\mathbf{log}(\mathbf{1}-\boldsymbol{\theta})) \\
\mathbf{0} &= \frac{\boldsymbol{n}}{\boldsymbol{\theta}} - \frac{\boldsymbol{N-n}}{\mathbf{1}-\boldsymbol{\theta}} \\
\boldsymbol{\theta} &= \frac{\boldsymbol{n}}{\boldsymbol{N}}
\end{aligned}
$$

Q6. In **Q5,** suppose that the prior for the probability $\boldsymbol{p}$ follows a Beta distribution

Find the **MAP** estimate of $\boldsymbol{\theta}$**.**

Solution

$$
\begin{aligned}
\boldsymbol{\theta_{MAP}} &= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\, \boldsymbol{p(x|\theta)p(\theta)} \\
&= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\ \left(\boldsymbol{\theta^{n}(1-\theta)^{N-n}\theta^{\alpha-1}(1-\theta)^{\beta-1}}\right) \\
&= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\ \left(\boldsymbol{\theta^{n+\alpha-1}(1-\theta)^{N-n+\beta-1}}\right) \\
&= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\, \boldsymbol{log}\left(\boldsymbol{\theta^{n+\alpha-1}(1-\theta)^{N-n+\beta-1}}\right) \\
&= \underset{\boldsymbol{\theta}}{\mathbf{argmax}}\ ((\boldsymbol{n+\alpha-1})\,\mathbf{log}\,\boldsymbol{\theta} + (\boldsymbol{N-n+\beta-1})\,\mathbf{log}(\mathbf{1}-\boldsymbol{\theta}))
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{0} &= \frac{\boldsymbol{\partial}}{\boldsymbol{\partial\theta}}((\boldsymbol{n+\alpha-1})\,\mathbf{log}\,\boldsymbol{\theta} + (\boldsymbol{N-n+\beta-1})\,\mathbf{log}(\mathbf{1}-\boldsymbol{\theta})) \\
\mathbf{0} &= \frac{\boldsymbol{n+\alpha-1}}{\boldsymbol{\theta}} - \frac{\boldsymbol{N-n+\beta-1}}{\mathbf{1}-\boldsymbol{\theta}} \\
\boldsymbol{\theta} &= \frac{\boldsymbol{\alpha}+\mathbf{n}-\mathbf{1}}{\mathbf{N}+\boldsymbol{\alpha}+\boldsymbol{\beta}-\mathbf{2}}
\end{aligned}
$$

Q7. Suppose that $\boldsymbol{X_1, \ldots, X_n}$ are independent identically distributed samples from a Poisson distribution. Each random variable $\boldsymbol{X_i}$ is distributed according to the Poission distribution whose **pdf** is given by:

$$\boldsymbol{p(X_i = k|\lambda)} = \frac{\boldsymbol{\lambda^{k}e^{-\lambda}}}{\boldsymbol{k}!}$$

where the parameter $\boldsymbol{\lambda}$ is known as the **mean** of the Poission distribution.

a. Determine the maximum likelihood estimate (**MLE**) of the **mean** $\boldsymbol{\lambda}$**.**
b. Work out the **pdf** of sum of the random variables $\boldsymbol{X_1 + \cdots + X_n}$

Solution

a.

$$\lambda_{MLE} = \underset{\lambda}{\mathrm{argmax}}\, p(X_1 = x_1, \dots, X_n = x_n|\lambda)$$

$$= \underset{\lambda}{\mathrm{argmax}} \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$= \underset{\lambda}{\mathrm{argmax}}\, log\left(\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right)$$

$$= \underset{\lambda}{\mathrm{argmax}} \sum_{i=1}^{n} \log \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$= \underset{\lambda}{\mathrm{argmax}} \sum_{i=1}^{n} (x_i \log \lambda - \lambda - \log(x_i!))$$

$$= \underset{\lambda}{\mathrm{argmax}}(-\lambda n + (\log \lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!))$$

$$0 = \frac{\partial}{\partial \lambda}\left(-\lambda n + (\log \lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!)\right)$$

$$0 = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i$$

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

b. For the sum of two poisson random variables we have:

$$p(X_1 + X_2 = k) = \sum_{i=0}^{k} p(X_1 + X_2 = k, X_1 = i)$$

$$= \sum_{i=0}^{k} p(X_2 = k - i, X_1 = i)$$

$$= \sum_{i=0}^{k} p(X_2 = k - i)\, p(X_1 = i)$$

$$= \sum_{i=0}^{k} \frac{e^{-\lambda} \lambda^{k-i}}{(k-i)!} \frac{e^{-\lambda} \lambda^{i}}{i!}$$

$$= e^{-(\lambda+\lambda)} \frac{1}{k!} \sum_{i=0}^{k} \frac{k!}{i!\,(k-i)!} \lambda^{k-i} \lambda^{i}$$

$$= e^{-(\lambda+\lambda)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^{k-i} \lambda^{i}$$
$$= e^{-(\lambda+\lambda)} \frac{(\lambda+\lambda)^k}{k!}$$

We can treat $X_1 + X_2$ as a new random variable and repeat the above process to add X_3. For the sum of n variables. We finally get

$$\mathbf{P}(\mathbf{X_1} + \cdots + \mathbf{X_n} = k) = \frac{\mathbf{e^{-n\lambda}(n\lambda)^k}}{\mathbf{k!}}$$

Q8. Two players **A** and **B** are competing at a trivia quiz game involving a series of questions. On any individual question, the probabilities that **A** and **B** give the correct answer are α and β respectively, for all questions, with outcomes for different questions being independent. The game finishes when a player wins by answering a question correctly.
Compute the probability that **A** wins if
(a) **A** answers the first question
(b) **B** answers the first question.

Solution

(a) The probability that **A** wins the first time is α.

To win on the second time:

1. **A** has to answer the question ***incorrectly***, then
2. **B** has to answer the question ***incorrectly***, then
3. **A** has to answer the question ***correctly***

The probability for:
1. is $(1-\alpha)$, for 2. the probability is $(1-\beta)$ and for 3. the probability is α.

Thus the probability that **A** wins on the second time is: $(1-\alpha)(1-\beta)\,\alpha$

In general, the probability that **A** wins on the $\mathbf{n^{th}}$ opportunity is:

$$p = ((1-\alpha)(1-\beta))^n \;\alpha, \qquad n \in [0, \infty]$$

The total probability that **A** wins the game is the sum of probabilities for all the above events:

$$p(A\,wins) = \sum_{n=0}^{\infty} ((1-\alpha)(1-\beta))^n \;\alpha$$
$$= \alpha \sum_{n=0}^{\infty} ((1-\alpha)(1-\beta))^n$$
$$= \frac{\alpha}{1-(1-\alpha)(1-\beta)}$$

by using the ***sum of infinite geometric series*:**

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \qquad |r| < 1$$

(b) The probability of events that **A** wins if **B** gives the first answer is:

$$\boldsymbol{p} = (\mathbf{1}-\boldsymbol{\beta})\ \left((\mathbf{1}-\boldsymbol{\alpha})(\mathbf{1}-\boldsymbol{\beta})\right)^{\boldsymbol{n}}\ \ \boldsymbol{\alpha}, \qquad \boldsymbol{n}\,\boldsymbol{\epsilon}\,[\mathbf{0},\infty]$$

In this case, the probability that **A** wins is

$$\begin{aligned}\boldsymbol{p}(\boldsymbol{A\,wins}) &= \sum_{\boldsymbol{n}=\mathbf{0}}^{\infty}(\mathbf{1}-\boldsymbol{\beta})\,\boldsymbol{\alpha}\,((\mathbf{1}-\boldsymbol{\alpha})(\mathbf{1}-\boldsymbol{\beta}))^{\boldsymbol{n}} \\ &= \boldsymbol{\alpha}(\mathbf{1}-\boldsymbol{\beta})\sum_{\boldsymbol{n}=\mathbf{0}}^{\infty}((\mathbf{1}-\boldsymbol{\alpha})(\mathbf{1}-\boldsymbol{\beta}))^{\boldsymbol{n}} \\ &= \frac{\boldsymbol{\alpha}\,(\mathbf{1}-\boldsymbol{\beta})}{\mathbf{1}-(\mathbf{1}-\boldsymbol{\alpha})(\mathbf{1}-\boldsymbol{\beta})}\end{aligned}$$

Q9. Patients are recruited onto the two arms (**0 - Control, 1 - Treatment**) of a clinical trial. The probability that an adverse outcome occurs on the control arm is $\boldsymbol{p_0}$ and is $\boldsymbol{p_1}$ for the treatment arm. Patients are allocated alternately onto the two arms, and their outcomes are independent.

What is the probability that the first adverse event occurs on the control arm?

Solution

As in Question 7, we can define the probability of the events that the first adverse event occurs on the control arm when patients start at the control arm as:

$$\boldsymbol{p} = ((\mathbf{1}-\boldsymbol{p_0})(\mathbf{1}-\boldsymbol{p_1}))^{\boldsymbol{n}}\,\boldsymbol{p_0} \qquad \boldsymbol{n}\,\boldsymbol{\epsilon}\,[\mathbf{0},\infty]$$

The probability we are looking for is the sum of probabilities for all the events:

$$\begin{aligned}\boldsymbol{p_{total}} &= \sum_{\boldsymbol{n}=\mathbf{0}}^{\infty}\boldsymbol{p_0}\ \ ((\mathbf{1}-\boldsymbol{p_0})(\mathbf{1}-\boldsymbol{p_1}))^{\boldsymbol{n}} \\ &= \frac{\boldsymbol{p_0}}{\mathbf{1}-(\mathbf{1}-\boldsymbol{p_0})(\mathbf{1}-\boldsymbol{p_1})}\end{aligned}$$

Similarly, if patients start at the treatment arm, then the probability that the first adverse event occurs on the control arm is given by:

$$\boldsymbol{p'} = (\mathbf{1}-\boldsymbol{p_1})\,((\mathbf{1}-\boldsymbol{p_0})(\mathbf{1}-\boldsymbol{p_1}))^{\boldsymbol{n}}\,\boldsymbol{p_0} \qquad \boldsymbol{n}\,\boldsymbol{\epsilon}\,[\mathbf{0},\infty]$$

The probability we are looking for is the sum of probabilities for all the events:

$$\begin{aligned}\boldsymbol{p'_{total}} &= \sum_{\boldsymbol{n}=\mathbf{0}}^{\infty}\boldsymbol{p_0}\,(\mathbf{1}-\boldsymbol{p_1})\,((\mathbf{1}-\boldsymbol{p_0})(\mathbf{1}-\boldsymbol{p_1}))^{\boldsymbol{n}} \\ &= \frac{\boldsymbol{p_0}(\mathbf{1}-\boldsymbol{p_1})}{\mathbf{1}-(\mathbf{1}-\boldsymbol{p_0})(\mathbf{1}-\boldsymbol{p_1})}\end{aligned}$$

Since, the question does not say whether patients started on the treatment arm or the control arm, we do not know which one of the above is correct.

On the other hand, if we were told that there is a probability $\boldsymbol{\theta}$ that patients start on the control arm then the probability that the first adverse event occurs on the control arm is given by:

$$\boldsymbol{\theta p_{total} + (1-\theta)\, p'_{total} = \frac{\theta p_0 + (1-\theta)\, p_0(1-p_1)}{1-(1-p_0)(1-p_1)}}$$