Review

- Consider random variables $X_1, X_2, \dots, X_n$, where $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$.

Define $W_n \triangleq a_1 X_1 + a_2 X_2 + \dots + a_n X_n$. Then,

$$E(W_n) = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n.$$

If X_i's are independent*:

$$Var(W_n) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2.$$

(* For the previous equation to be correct, it's enough that X_i's are uncorrelated.)

Special case

Assume that X_i's are i.i.d. (independent and identically distributed). Hence,

$$\forall i: \begin{cases} \mu_i = \mu, \\ \sigma_i^2 = \sigma^2. \end{cases}$$

Let $a_i = \frac{1}{n}$ in W_n, denote the resulting random variable by $\overline{X}_n$; i.e.,

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Then,

$$E(\overline{X}_n) = n \cdot \frac{\mu}{n} = \mu,$$

$$Var(\overline{X}_n) = n \cdot \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

* Notice that $Var(\overline{X}_n)$ decreases as n increases.

Law of Large Numbers

- For any positive ϵ,

$$\lim_{n\to\infty} \Pr(|\bar{X}_n-\mu|>\epsilon)=0.$$

Proof. To prove the result, we use Chebyshev inequality that for a random variable Y with $E(Y)=\mu_y$ and $Var(Y)=\sigma_y^2$,*

$$P(|Y-\mu_y|\geq\epsilon)\leq\frac{\sigma_y^2}{\epsilon^2}.$$

Now, using Chebyshev inequality for $\bar{X}_n$, we get:

$$P(|\bar{X}_n-\mu|\geq\epsilon)\leq\frac{\sigma^2/n}{\epsilon^2}=\frac{\sigma^2}{n\epsilon^2}.$$

Hence,

$$0\leq\lim_{n\to\infty}(|\bar{X}_n-\mu|\geq\epsilon)\leq\lim_{n\to\infty}\frac{\sigma^2}{n\epsilon^2}=0\Rightarrow\lim_{n\to\infty}(|\bar{X}_n-\mu|\geq\epsilon)=0.$$ ■

(* Here is a proof of Chebyshev inequality:

$$\sigma_y^2=\int_{-\infty}^{+\infty}(y-\mu_y)^2f(y)\,dy\geq\int_{|y-\mu_y|\geq\epsilon}(y-\mu_y)^2f(y)\,dy\geq\epsilon^2\int_{|y-\mu_y|\geq\epsilon}f(y)\,dy=\epsilon^2P(|Y-\mu_y|\geq\epsilon)$$

$$\Rightarrow P(|Y-\mu_y|\geq\epsilon)\leq\frac{\sigma_y^2}{\epsilon^2}.)$$

More About the Law of Large Numbers (LLN)

The LLN justifies our intuition that the probability of a repeatable event A can be estimated by the relative frequency of its occurrence for a large number of independent repetitions. (Hence, it fills the gap between theory and the real world!)

To see this, assume that we want to estimate the probability of event A in practice. For that, introduce random variable

$$X_i = \begin{cases} 1 & \text{if } A \text{ occurs at the } i\text{-th trial} \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$E(X_i) = P(X_i = 1) \times 1 + P(X_i = 0) \times 0 = P(X_i = 1)$$

Denote $P(X_i = 1)$ by p. Using the LLN:

$$\Pr\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \overbrace{E\left(\frac{X_1 + \dots + X_n}{n}\right)}^{=p}\right| < \epsilon\right) \xrightarrow{n \to \infty} 1$$

$$\Rightarrow \Pr\left(\left|\frac{X_1 + \dots + X_n}{n} - p\right| < \epsilon\right) \xrightarrow{n \to \infty} 1$$

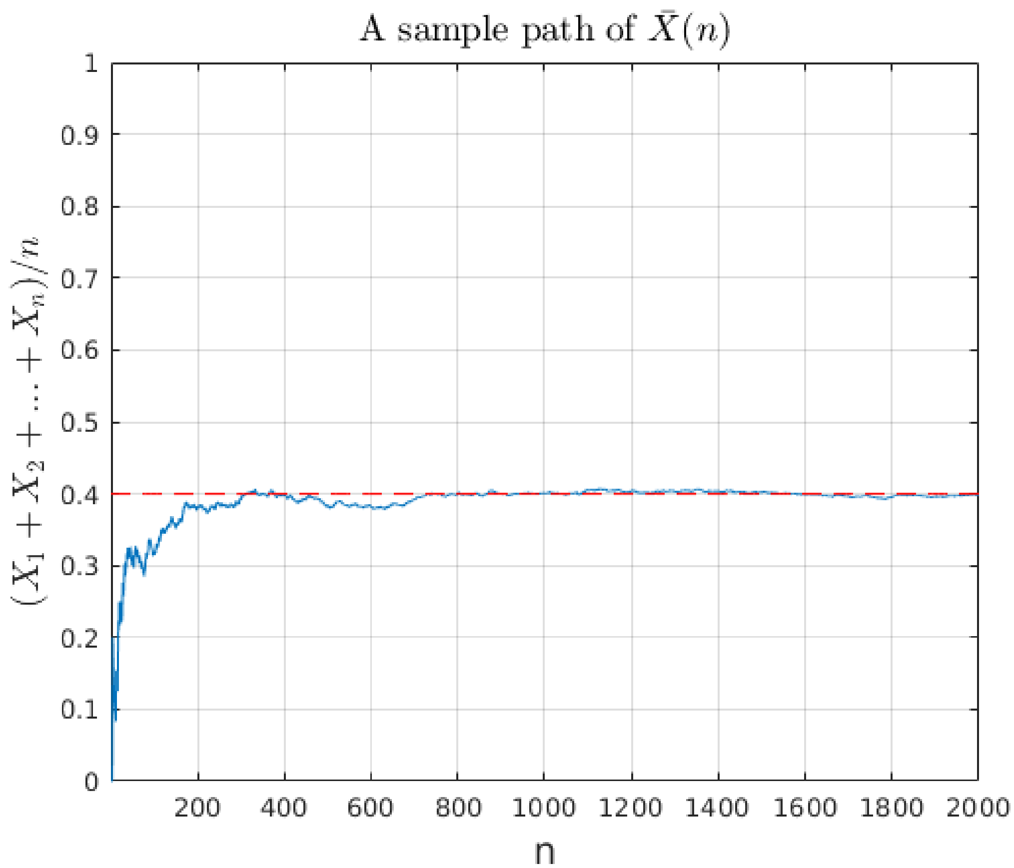
This means that if the number of successes of A in n trials equals k, then with probability one, $\frac{k}{n} \to p$ as $n \to \infty$.

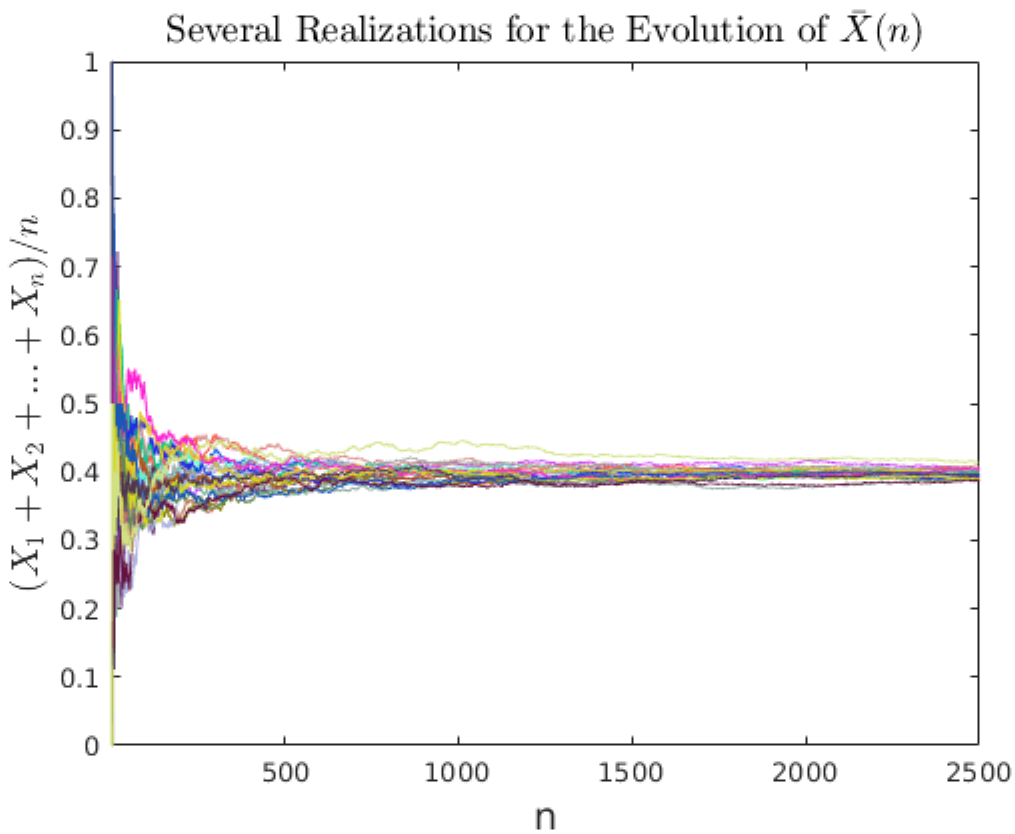
As an example, consider a coin in which we do not know P(Head) for it. We perform a large number of trials (tossing the coin) and see how many times it shows "Head". The relative frequency of "Head"s gives us an estimate

$p = P(\text{Head})$.

```
clear,close all
p = 0.4; %% set the probability of "head" to an arbitrary number in
 [0,1]

% Simulation of a sample path for evolution of X_bar
N = 2000;    %% number of trials

S = zeros(1,N+1); %% initiate the number of successes vector
X_bar = zeros(1,N+1); %% initiate the success frequency vector
for n = 1 : N
    X_n = (rand(1) < p);   %% (just a way to) simulate a coint toss
    S(n+1) = S(n) + X_n;   %% update the number of success vector
    X_bar(n+1) = S(n+1)/n; %% update the success frequency vector

    % plot the results
    if mod(n,10) == 0
        figure(1)
        plot(1:n,X_bar(2:n+1)) %% plot the success frequency vs number
 of trials
        xlim([1,N])          %% set limits for x-axis
        ylim([0,1])
        xlabel({'n'},'FontSize',14);
        ylabel({'$(X_1+X_2+...+X_n)/n
$'},'Interpreter','latex','FontSize',14)
        grid on
        pause(.001)            %% pause so we can see the evolution
 of the output graph
    end
end
hold on
plot(1:N,p*ones(1,N),'r--') % plot the line for the actual p used
title({'A sample path of $\bar{X}
(n)$'},'Interpreter','latex','FontSize',14)
pause(2)

% Sample paths of X_bar for several different realizations
figure(2)
N = 2500;
for m = 1:25 % iterate for different realizations
    S = zeros(1,N+1); %% initiate the number of successes vector
    X_bar = zeros(1,N+1); %% initiate the success frequency vector
    for n = 1 : N
        X_n = (rand(1) < p);   %% (just a way to) simulate a coint
 toss
        S(n+1) = S(n) + X_n;   %% update the number of success vector
        X_bar(n+1) = S(n+1)/n; %% update the success frequency vector
    end
    % plot the results
    figure(2)
    plot(1:N,X_bar(2:end),'Color', rand(1,3)) %% plot the success
 frequency vs number of trials
    hold on
```

```
    pause(0.025)
end
xlim([1,N])             %% set limits for x-axis
ylim([0,1])
xlabel({'n'},'FontSize',14);
ylabel({'$(X_1+X_2+...+X_n)/n$'},'Interpreter','latex','FontSize',14)
title({'Several Realizations for the Evolution of $\bar{X}
(n)$'},'Interpreter','latex','FontSize',14)
```





Published with MATLAB® R2016b

Problem A.12

a) We solve this problem for $N=2$ and $p=0.8$. (Instead of $N=2$ and $p=0.2$)

$$P(\text{correct decoding}) = P(\text{majority of identical digits decoded correctly on their own})$$

$$= \sum_{i=N+1}^{2N+1} \binom{2N+1}{i} p^i (1-p)^{2N+1-i}.$$

For $N=2$ and $p=0.8$: $P(\text{correct decoding}) = 0.942$.

b) $2N+1=3 \Rightarrow N=1$.

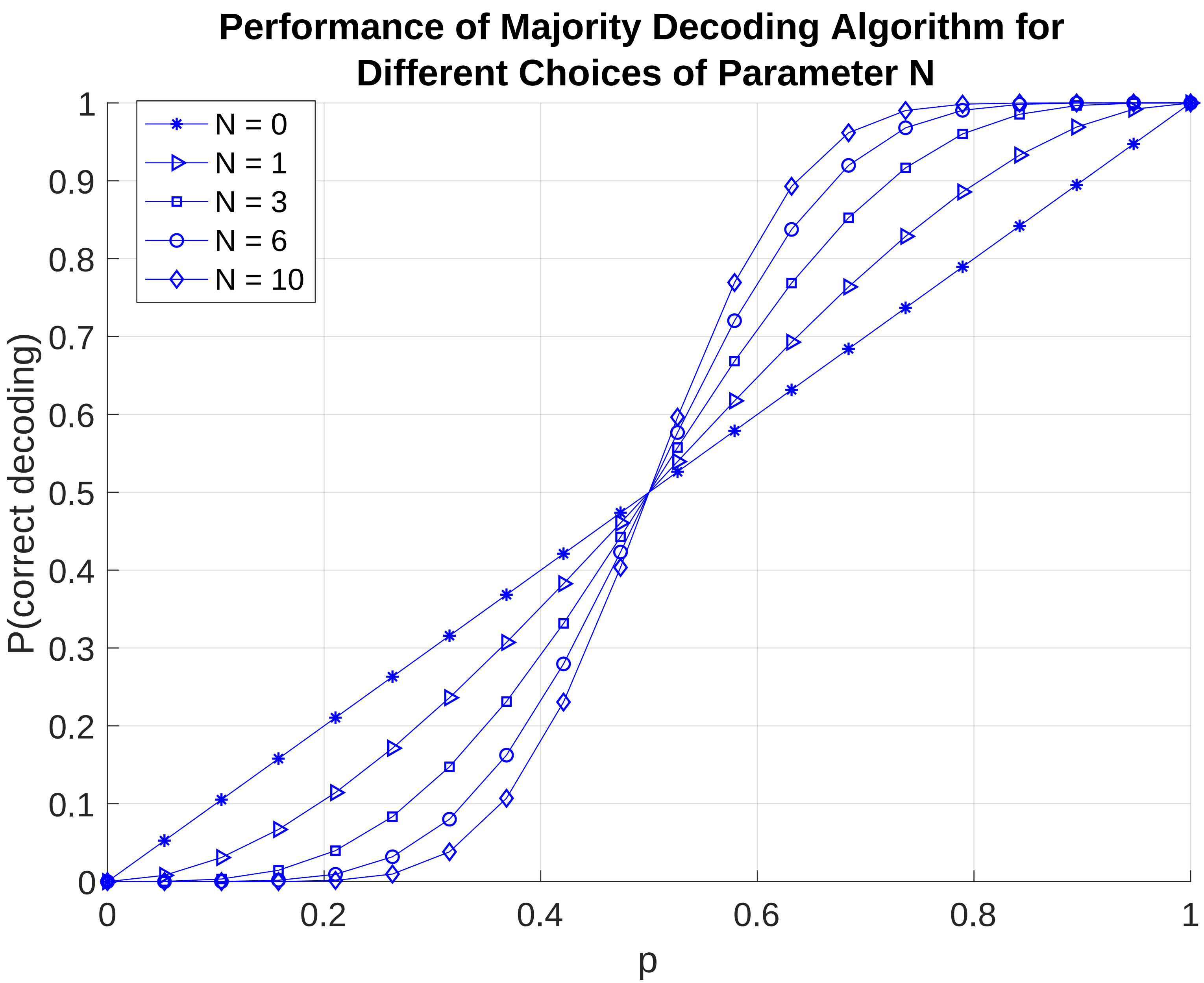
$$\sum_{i=2}^{3} \binom{3}{i} p^i (1-p)^{3-i} \geqslant 0.942 \xrightarrow[p]{\text{solve for}} p \geqslant 0.8537.$$

So, to get the same performance with adding less redundancy, we must have a better channel (higher p).

c) $2N+1=7 \Rightarrow N=3$.

$$\sum_{i=4}^{7} \binom{7}{i} p^i (1-p)^{7-i} \geqslant 0.942 \Longrightarrow p \geqslant 0.7645.$$

Hence, if we use a worse channel, we need to add more redundancy to get the same performance.


Performance of Majority Decoding Algorithm for
Different Choices of Parameter N
N = 0
N = 1
N = 3
N = 6
N = 10
P(correct decoding)
p
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

Problem A.14

For Binomial distribution:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \leq p \leq 1$$

To find p^* that maximizes $P(X=k)$, we differentiate $P(X=k)$ w.r.t. p.

Since $P(X=k)$ and $\ln(P(X=k))$ have their maximum at the same poin

we find p^* by setting the derivative of $\ln P(X=k)$ equal to 0:

$$\frac{\partial}{\partial p} \ln P(X=k) = \partial(k \ln p + (n-k)\ln(1-p))/\partial p = \frac{k}{p} - \frac{n-k}{1-p} = 0 \Rightarrow \boxed{p^* = \frac{k}{n}, 0<k} \; (1)$$

$$\left[\frac{\partial^2 \ln P(X=k)}{\partial p^2} = -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} \leq 0 \Rightarrow p^* \text{ is maximizer}\right]$$

For $k=0$, $P(X=k) = (1-p)^n \Rightarrow \boxed{p^*=0}$ ($(1-p)^n$ decreasing in $[0,1]$) (2)

For $k=n$, $P(X=k) = p^n \Rightarrow \boxed{p^*=1}$ (p^n increasing in $[0,1]$) (3)

(1), (2), (3) $\Rightarrow \boxed{p^* = \frac{k}{n}, \quad k=0,1,\dots,n}$

* What we did here is the Maximum Likelihood estimation of paramet

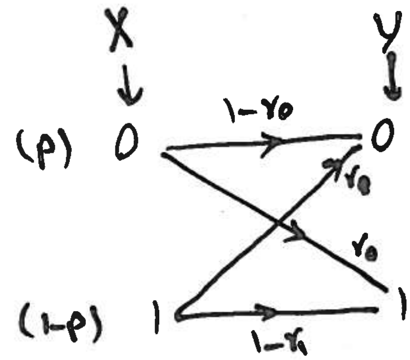
p, given an observation of the model.

Maximum Likelihood is a method to estimate the parameters of a statistica

model given observations. In particular, it chooses the parameter that most

likely caused the observed data to occure.

Problem A.17



a) $P(\text{correct decoding}) = P(Y=1, X=1) + P(Y=0, X=0)$

$= P(Y=1 \mid X=1)P(X=1) + P(Y=0 \mid X=0)P(X=0)$

$= (1-r_1)(1-p) + (1-r_0)p$

b) $P(Y_1=1, Y_2=0, Y_3=1, Y_4=1 \mid X_1=1, X_2=0, X_3=1, X_4=1)$

$= P(Y_1=1 \mid X_1=1, X_2=0, X_3=1, X_4=1)\,P(Y_2=0 \mid X_1=1, X_2=0, X_3=1, X_4=1)$

$P(Y_3=1 \mid X_1=1, X_2=0, X_3=1, X_4=1)\,P(Y_4=1 \mid X_1=1, X_2=0, X_3=1, X_4=1)$

$= P(Y_1=1 \mid X_1=1)P(Y_2=0 \mid X_2=0)P(Y_3=1 \mid X_3=1)P(Y_4=1 \mid X_4=1)$

$= (1-r_1)^3(1-r_0)$

c) $(X_1, X_2, X_3) = (0,0,0)$ transmitted

Correct decoding when all Y_1, Y_2, Y_3 are '0' or one of them is '1' and the two others are '0':

$P_{maj}(\text{correct decoding}) = P(Y_1, Y_2, Y_3 \text{ are all '0'} \mid X_1=X_2=X_3=0)$

$+ P(\text{one of } Y_1, Y_2, Y_3 \text{ is '1' \& others are '0'} \mid X_1=X_2=X_3=0)$

$= (1-r_0)^3 + \binom{3}{1} r_1 (1-r_0)^2$

d) Find r_0 such that $(1-r_0)^3 + 3r_1(1-r_0)^2 \geq (1-r_0)$:

$\Longleftrightarrow$ $\xrightarrow{\text{since } 0 \le r_0 \le 1}$ $0 \leq r_0 \leq 1 - \frac{1}{2}\left(\sqrt{9r_1^2+4} - 3r_1\right)$

$\begin{cases} r_0 \leq 1 + \left(+\frac{3}{2}r_1 - \frac{1}{2}\sqrt{9r_1^2+4}\right) \\ \ldots \leq r_0 \leq 1 + \left(\frac{3}{2}r_1 + \frac{1}{2}\sqrt{9r_1^2+4}\right) \end{cases}$

Problem A.17 (Cont'd)

e) ~~Let~~ $P(X_1=0, X_2=0, X_3=0 \mid Y_1=1, Y_2=0, Y_3=1)$

$$= \frac{P(Y_1=1, Y_2=0, Y_3=1 \mid X_1=0, X_2=0, X_3=0)\,P(X_1=0, X_2=0, X_3=0)}{P(Y_1=1, Y_2=0, Y_3=1)}$$

$$= \frac{r_0^2(1-r_0)p}{P(Y_1=1, Y_2=0, Y_3=1)} \quad ①$$

Since, there are only two options for the transmitted sequence:

$$P(Y_1=1, Y_2=0, Y_3=1) = P(Y_1=1, Y_2=0, Y_3=1 \mid X_1=0, X_2=0, X_3=0)\,P(X_1=0, X_2=0, X_3=0)$$
$$+ P(Y_1=1, Y_2=0, Y_3=1 \mid X_1=1, X_2=1, X_3=1)\,P(X_1=1, X_2=1, X_3=1)$$
$$= r_0^2(1-r_0)p + (1-r_1)^2 r_1 (1-p) \quad ②$$

①, ② $\Rightarrow$ $P(X_1=0, X_2=0, X_3=0 \mid Y_1=1, Y_2=0, Y_3=1) = \boxed{\dfrac{r_0^2(1-r_0)p}{r_0^2(1-r_0)p + (1-r_1)^2 r_1 (1-p)}}$