

Textbook Chapter 3

3.1

3.2

3.3

Odd Questions

Question 5

Question 6

Textbook Chapter 3

3.1

The univariate probability density function $p(d) = c(1 - d)$ is defined on the interval $0 \leq d \leq 1$.

(A) What must the constant, c , be for the probability density function to be properly normalized?

(B) Calculate the mode, median, and mean of this probability density function analytically.

Answer:

(A) To normalize $p(d)$, we will have

$$\begin{aligned}\int_0^1 p(d) dd &= \int_0^1 c(1 - d) dd = 1 \\ c(d - \frac{1}{2}d^2) \Big|_0^1 &= 1 \\ c &= 2\end{aligned}$$

Thus, $c = 2$.

(B)

i). The mean value $\mu = \int dp(d)dd$, with $p(d) = 2(1 - d)$, we will have

$$\mu = \int_0^1 2d(1 - d)dd = \int_0^1 (2d - 2d^2)dd = d^2 - \frac{2}{3}d^3 \Big|_0^1 = \frac{1}{3}$$

ii). The median value m should satisfy: $\int_0^m p(d)dd = \frac{1}{2}$

$$\begin{aligned}\int_0^m 2(1 - d)dd &= \frac{1}{2} \\ (2d - d^2) \Big|_0^m &= \frac{1}{2} \\ 2m - m^2 &= \frac{1}{2} \\ (m - 1)^2 &= \frac{1}{2}\end{aligned}$$

Thus, there are two possible solution for m , $m = 1 \pm \frac{\sqrt{2}}{2}$

with the given condition $0 \leq d \leq 1$, we will have the median is

$$m = 1 - \frac{\sqrt{2}}{2}$$

iii). The mode M should satisfy: $p(M) \geq p(d)$,

The derivative of $p(d) = 2(1 - d)$ is $p'(d) = -2$, which means $p(d)$ continuously decreases with the increase of d , and the maximum value of the function occur at $d = 0$

So, $M = 0$.

In summary, the mean value is $\frac{1}{3}$, the median value is $1 - \frac{\sqrt{2}}{2}$, the mode is 0.

3.2

The univariate exponential probability density function is

$p(d) = \lambda \exp(-\lambda d)$ where d is defined on the interval $0 \leq d < \infty$. The parameter, λ , is called the rate parameter.

(A) Use MatLab to plot shaded column-vectors of this probability density function and to compute its mode, median, mean, and variance, for the cases $\lambda = 5$ and $\lambda = 10$.

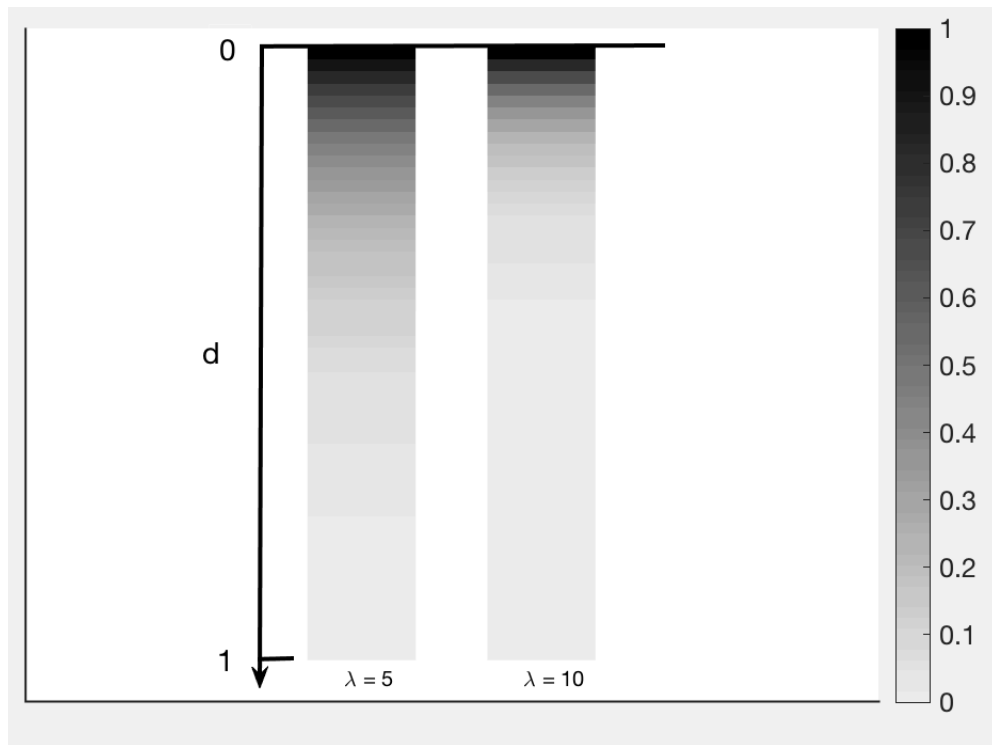
(B) Is it possible to control the mean and variance of this probability density function separately?

Answer:

(A).

The probability density function is $p(d) = \lambda \exp(-\lambda d)$,

Answer:



(The arrow axis is drawn by iPad on the Matlab output image)

The raw output:

```
ans =
```

```
"when  $\lambda$  = 5, mean = 0.200, median = 0.139,  
variance = 0.040, mode = 0.000"
```

```
ans =
```

```
"when  $\lambda$  = 10, mean = 0.100, median = 0.069,  
variance = 0.010, mode = 0.000"
```

which in summary is

$\lambda =$	5	10
Mean	0.2	0.1
Median	0.139	0.069
Variance	0.04	0.01
Mode	0	0

The script is:

```
clearvars; % clean up workspace

% plot part
% to better demonstrate the color gradient tendency,
% d is chosen to be  $0 \leq d \leq 1$ , as  $p(1) = 5 \exp(-5)$  or
%  $p(1) = 10 \exp(-10)$  are both approximate to 0

% parameter settings
d_min = 0;
d_max = 1;
n = 50;
interval = d_max / n;
d = [d_min:interval:d_max]';

% plot initialization
figure(2);
clf;
set(gca, 'LineWidth', 1, 'FontSize', 14);

axis equal;
set(gca, 'XTick', []); % turn off horizontal axis
set(gca, 'YTick', []); % turn off vertical axis
```

```

% ? borrowed from 'eda/ch03/eda03_06.m'

colormap(flipud(gray)*0.93);
hold on;

% ? color setting, with help of
https://www.mathworks.com/help/matlab/ref/colormap.html

for l=[5 10]
    p = l*exp(-l*d);
    range = max(p)-min(p);
    imagesc( [3*l,3*l+3], [n , 0], (p-min(p))/range);
    text(3*l,-2,sprintf('\lambda = %d',l));
    hold on;
end

colorbar;

clearvars;

% calculation part
% to calculate it precious, as  $0 \leq d < \infty$ ,
% d_max is set to be 10000

% parameter settings
d_min = 0;
d_max = 10000;
n = 1000000;
interval = d_max / n;
d = [d_min:interval:d_max]';

for l=[5 10]

```

```

p = @(d) 1*exp(-1*d);

% mean = E(x) = \int x * p(x) dx

p_d = @(d) p(d).*d;
meanval = integral( p_d,0,Inf);

% variance = E(x^2) - E(x)^2
%           = \int x^2 * p(x) dx - mean ^2

p_d2 = @(d) p(d).*d.^2;
variance = integral(p_d2,0,Inf) - meanval^2;

% median M: \int_0^M p(x) dx - 1/2 = 0

intp = @(d) integral (p,0,d) - 0.5;
medianval = fzero(intp,0);

% mode is the d of p(d) = p_max

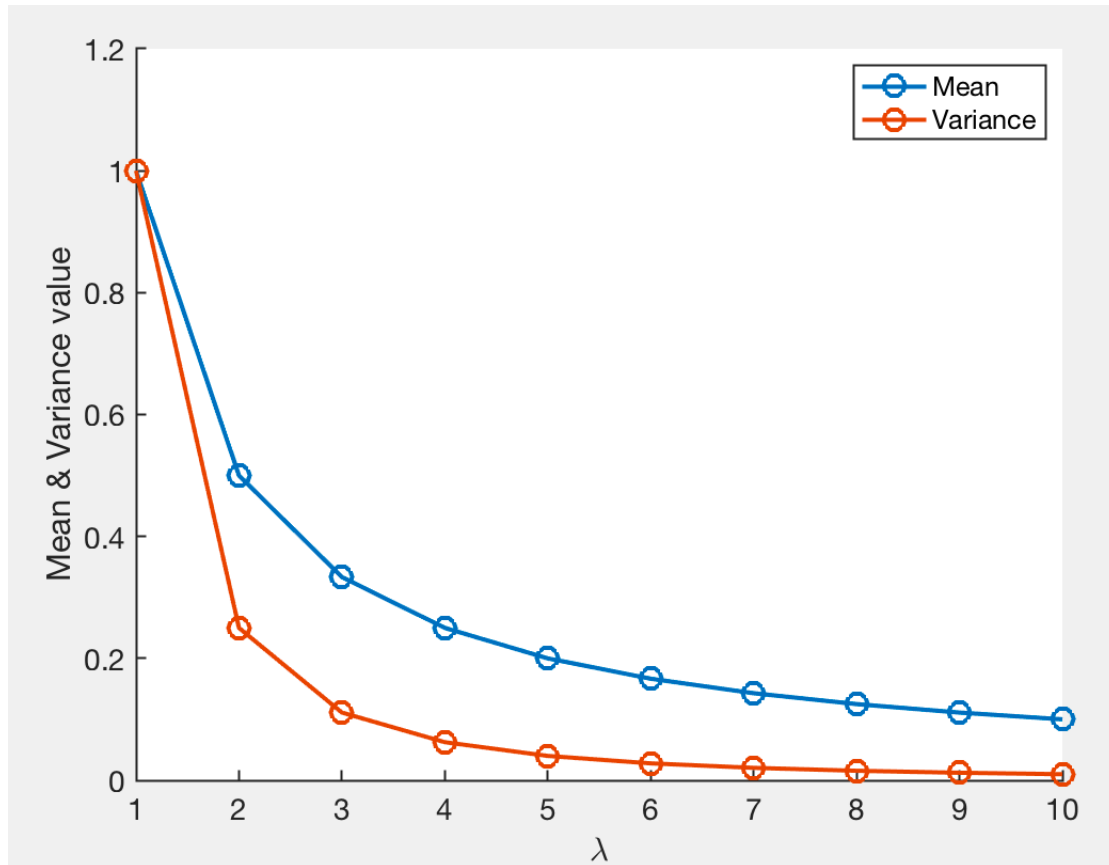
[val,index] = max(p(d));
modeval = d(index);

sprintf("when $\lambda$ = %d, mean = %.3f, median =
%.3f, variance = %.3f, mode = %.3f",1,meanval,medianval,
variance, modeval)
end

```

(B).

To test the dependence, here are mean and variance table with the parameter λ from 1 to 10. It is shown that mean values and variance values both vary with the λ . So we cannot control the mean and variance of this probability density function separately. The analytical process is followed.



The mean is

$$\begin{aligned}
 \mu &= E(D) = \int dp(d)dd = \int_0^{\infty} \lambda d \exp(-\lambda d) dd \\
 &= \left[-\frac{(\lambda d + 1) \exp(-\lambda d)}{\lambda} \right]_0^{\infty} \\
 &= \frac{1}{\lambda}
 \end{aligned}$$

The variance is

$$\begin{aligned}
 \sigma^2 &= E(D^2) - E(D)^2 = \int_0^{\infty} \lambda d^2 \exp(-\lambda d) dd - \frac{1}{\lambda^2} \\
 &= \left[-\frac{(\lambda^2 d^2 + 2\lambda d + 2) \exp(-\lambda d)}{\lambda} \right]_0^{\infty} - \frac{1}{\lambda^2} \\
 &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
 \end{aligned}$$

Thus, variance and mean are both related to λ , and $\text{variance} = \text{mean}^2$.
So They cannot be controlled separately.

Script for the plot:

```

clearvars;

% plot initialization
figure(2);
clf;
set(gca, 'LineWidth', 1, 'FontSize', 14);
hold on;

% parameter settings
l_list = 1:1:10;

meanlist = zeros(10,1);
varlist = zeros(10,1);

for l=l_list
    p = @(d) l*exp(-l*d);

    % mean = E(x) = \int x * p(x) dx

    p_d = @(d) l*exp(-l*d) .* d;
    meanlist(l) = integral( p_d, 0, Inf);

```

```

% variance = E(x^2) - E(x)^2
%           = \int x^2 * p(x) dx - mean ^2

p_d2 = @(d) 1*exp(-1*d) .* d .*d;
varlist(l) = integral(p_d2,0,Inf) - meanlist(l)^2;

end

plot(l_list,meanlist,"-o",l_list,varlist,"-o",...
      'LineWidth',2, 'MarkerSize',10);
legend('Mean','Variance');

xlabel(sprintf('\\lambda'));
ylabel('Mean & Variance value');

```

3.3

Suppose that $p(d)$ is a Normal probability density function with zero mean and unit variance.

(A) Derive the probability density function of $m = |d|^{1/2}$, analytically.

(B) Use MatLab to plot shaded column-vectors of this probability density function and to compute its mode, median, mean, and variance.

Answer:

(A). $p(m) = p[d(m)] \left| \frac{\partial d}{\partial m} \right|$

As $p(d)$ is a Normal probability density function with zero mean and unit variance, we have

$$p(d) = N(0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{d^2}{2}\right)$$

Since $m = |d|^{1/2}$, we can get $|d(m)| = m^2$ and $|\frac{\partial d}{\partial m}| = |2m|$.

$$\begin{aligned} p(m) &= \frac{2m}{\sqrt{2\pi}} \exp\left(-\frac{m^4}{2}\right) \\ &= \sqrt{\frac{2}{\pi}} m \exp\left(-\frac{m^4}{2}\right), \quad (m \geq 0) \end{aligned}$$

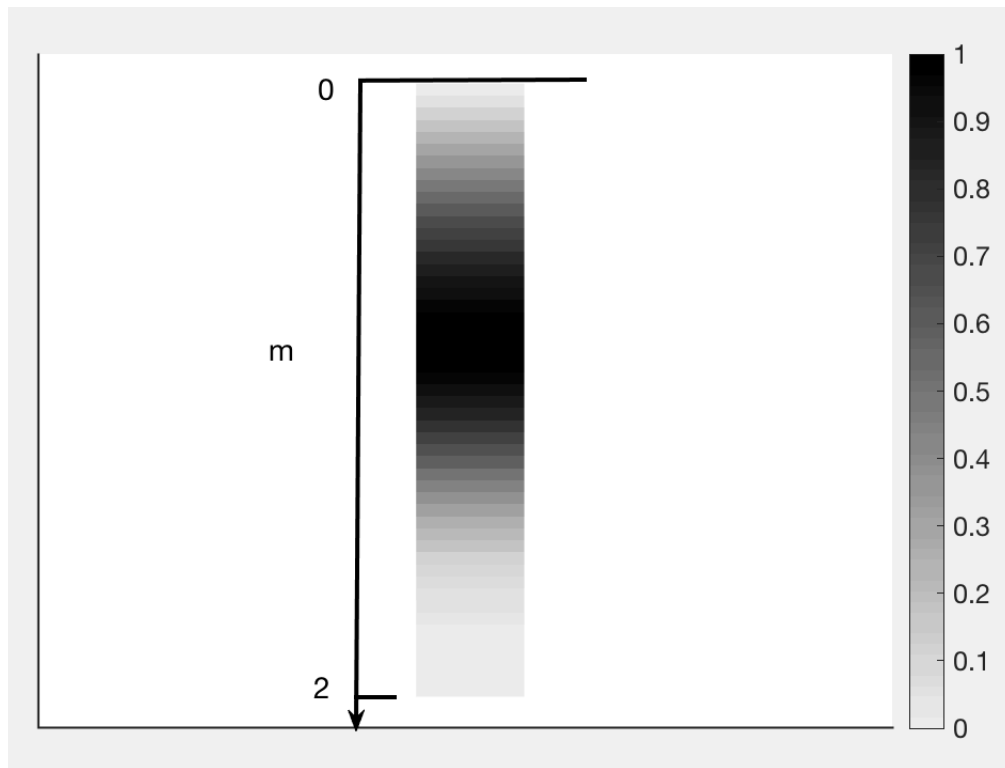
However this $p(m)$ is not normalized, thus, we should have

$$\int_0^\infty (p_{\text{normalized}}(m)) dm = \int_0^\infty (\text{const} \cdot p(m)) dm = 1$$

Thus, $\text{const} = 2$

$$p_{\text{normalized}}(m) = 2\sqrt{\frac{2}{\pi}} m \exp\left(-\frac{m^4}{2}\right), \quad (m \geq 0)$$

(B).



The raw output is

```
ans =  
  
    "mean = 0.822, median = 0.821, variance = 0.122, mode =  
    0.841"
```

The script is:

```
clearvars; % clean up workspace  
  
% plot part  
% to better demonstrate the color gradient tendency,  
% m_max is set to be 2, cuz p(2) appximates 0
```

```

% parameter settings
m_min = 0;
m_max = 2;
n = 50;
interval = m_max / n;
m = [m_min:interval:m_max]';

% plot initialization
figure(2);
clf;
set(gca, 'LineWidth', 1, 'FontSize', 14);

axis equal;
set(gca, 'XTick', []); % turn off horizontal axis
set(gca, 'YTick', []); % turn off vertical axis

% ? borrowed from 'eda/ch03/eda03_06.m'

colormap(flipud(gray)*0.93);
hold on;

% ? color setting, with help of
https://www.mathworks.com/help/matlab/ref/colormap.html

p = 2 * 2.*m/(sqrt(2*pi)) .* exp(- m.^4 / 2) ;
range = max(p)-min(p);
imagesc( [0,3], [n , 0], (p-min(p))/range);
colorbar;
hold on;

% Calculate part

clearvars;

```

```

p = @(m) 2*2.*m/(sqrt(2*pi)) .* exp(- m.^4 /2) ;

% mean = E(x) = \int x * p(x) dx

p_m = @(m) p(m).*m;
meanval = integral( p_m,0,Inf);

% variance = E(x^2) - E(x)^2
%           = \int x^2 * p(x) dx - mean ^2

p_m2 =@(m) p(m).*m.^2;
variance = integral(p_m2,0,Inf) - meanval^2;

% median M: \int_0^M p(x) dx - 1/2 * \int_0^Inf p(x) dx = 0

intp = @(m) 2*integral(p,0,m) - integral(p,0,Inf);
medianval = fzero(intp, [0,10]);

% mode is the d of p(d) = p_max

m_min = 0;
m_max = 2;
n = 5000;
interval = m_max / n;
m = [m_min:interval:m_max]';

[val,index] = max(p(m));
modeval = m(index);

sprintf("mean = %.3f, median = %.3f, variance = %.3f, mode =
%.3f",meanval,medianval, variance, modeval)

```

Odd Questions

Question 5

Assume the occurrence of Green Cabs is G , Blue Cabs is B ,

We know: $P(G) = 0.85$, $P(B) = 0.15$

Witness correctly reported the blue is W , $P(W|B) = 0.8$

Accordingly $P(W|G) = 0.2$

Thus,

$$P(W) = P(B)P(W|B) + P(G)P(W|G) = 0.15 \cdot 0.8 + 0.85 \cdot 0.2 = 0.29$$

To calculate the probability that sidesiper was blue, we shall calculate $P(B|W)$

$$P(B|W) = \frac{P(B)P(W|B)}{P(W)} = \frac{0.15 \cdot 0.8}{0.29} = 0.413793103$$

$$P(G|W) = 1 - P(B|W) \approx 0.586 > P(B|W)$$

Since $P(G|W) > P(B|W)$, It is more likely that sideswiper was green.

So (d) is correct!

Question 6

The patient has strep throat is S , and then assume the probability that doctor thinks the patient have strep throat $P(S) = s$

The test says YES is T_{yes} , and the test says NO is T_{no} ,

$$P(T_{yes}) + P(T_{no}) = 1$$

So, $P(T_{yes}|S) = 0.7, P(T_{no}|S) = 0.3, P(T_{yes}|\bar{S}) = 0.1, P(T_{no}|\bar{S}) = 0.9,$

The results are YES, NO, YES, NO, YES: $T_{yes}T_{no}T_{yes}T_{no}T_{yes},$

The results are independent with each other, so

$$P(T_i T_j | S) = P(T_i | S) P(T_j | S)$$

If the patient got the results in the condition of having strep throat, the probability is

$$\begin{aligned} P(T_{yes}T_{no}T_{yes}T_{no}T_{yes}|S) &= P(T_{yes}|S)P(T_{no}|S)P(T_{yes}|S)P(T_{no}|S)P(T_{yes}|S) \\ &= 0.7 \cdot 0.3 \cdot 0.7 \cdot 0.3 \cdot 0.7 \\ &= 0.03087 \end{aligned}$$

If the patient got the results in the condition of not having strep throat, the probability is

$$\begin{aligned} P(T_{yes}T_{no}T_{yes}T_{no}T_{yes}|\bar{S}) &= P(T_{yes}|\bar{S})P(T_{no}|\bar{S})P(T_{yes}|\bar{S})P(T_{no}|\bar{S})P(T_{yes}|\bar{S}) \\ &= 0.1 \cdot 0.9 \cdot 0.1 \cdot 0.9 \cdot 0.1 \\ &= 0.00081 \end{aligned}$$

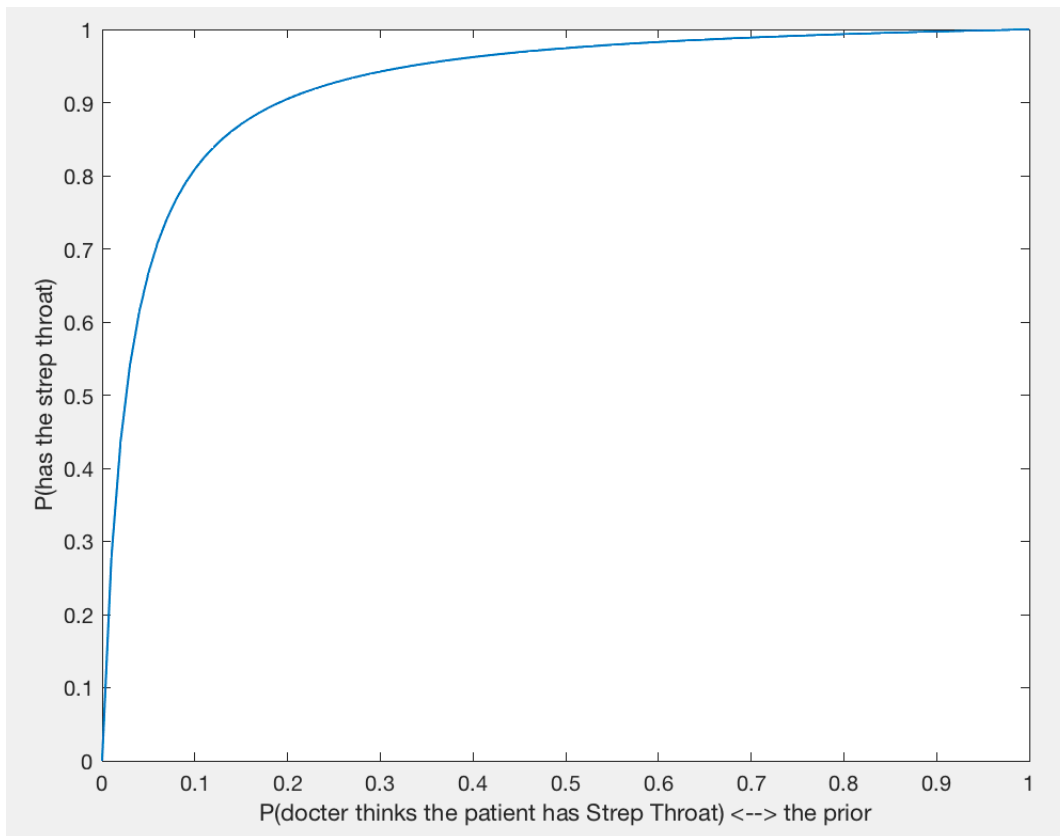
So, probability for the patient with these results to have strep throat is

$P(S|T_{yes}T_{no}T_{yes}T_{no}T_{yes})$, by applying Bayes's rule:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})}$$

$$\begin{aligned} P(S|T_{yes}T_{no}T_{yes}T_{no}T_{yes}) &= \frac{P(S)P(T_{yes}T_{no}T_{yes}T_{no}T_{yes}|S)}{P(S)P(T_{yes}T_{no}T_{yes}T_{no}T_{yes}|S) + P(\bar{S})P(T_{yes}T_{no}T_{yes}T_{no}T_{yes}|\bar{S})} \\ &= \frac{s \cdot 0.03087}{s \cdot 0.03087 + (1 - s) \cdot 0.00081} \end{aligned}$$

Then, I generate the plot of this probability with respect of s ,



If the doctor's assumption for a patient to have strep throat is higher than 0.5, the calculated value would be higher than 0.9. So, it is more likely that the patient has strep throat!

(d) is correct!

Script:

```
clearvars;
s = [0:0.01:1]';
p = 0.03087.*s ./ (0.03087.*s + (1-s)*0.00081);
plot(s,p,...
      'LineWidth', 1);

xlabel(sprintf('P(docter thinks the patient has Strep Throat)
<--> the prior'));
ylabel('P(has the strep throat)');
```