# CS 530 Geometric and Prob. Methods in CS
# Homework 1

Erik Barry Erhardt

August 30, 2006

Note: All code can be found in the Appendix.

**Exercise 1** *Independence.*

Variables $X$ and $Y$ will be independent if the joint probability is the product of the marginals for every pair of $x$ and $y$. The joint pmf is

```
5/12        1/18        5/72
1/36        1/36        5/36
1/12        1/6         1/72
```

and the joint pmf assuming $X$ and $Y$ are independent (product of marginals) is

```
247/864     13/96       13/108
133/1296     7/144       7/162
361/2592    19/288      19/324
```

Because these two distributions are not identical, $X$ and $Y$ are not independent.

**Exercise 2** *Marginal and conditional.*

**(a)** $p_X(x_i)$

```
4/9
5/18
5/18
```

**(b)** $p_Y(y_i)$

```
1/12          1/2           5/12
```

**(c)** $p_{X|Y}(x_i|y_i)$

```
  0           2/9           4/5
2/3           5/18          1/5
1/3           1/2            0
```

**(d)** $p_{Y|X}(y_i|x_i)$

```
  0           1/4           3/4
1/5           1/2           3/10
1/10          9/10           0
```

**Exercise 3** *Choose.*

Both choose functions reduce to the same expression of factorials, therefore they are the same.

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$\binom{n}{n-r} = \frac{n!}{(n-r)!(n-(n-r))!} = \frac{n!}{(n-r)!r!} = \frac{n!}{r!(n-r)!}$$

**Exercise 4** *Binomial stoplight.*
Problem setup:

$$\begin{aligned} X &= \text{Number of times stopped in a week (10 trials)} \\ \Pr(\text{stop}) &= 0.27 \\ X &\sim \text{Bin}(n = 10, p = 0.27) \\ \Pr(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \end{aligned}$$

Solution:

$$\begin{aligned} \Pr(X \geq 5) &= \sum_{x=5}^{10} \binom{10}{x} 0.27^x (1-0.27)^{10-x} \\ &= 0.074961 + 0.023104 + 0.0048831 + 0.00067728 + 0.000055667 + 0.0000020589 \\ &= 0.10368 \end{aligned}$$

**Exercise 5** *pmf: c and first two moments.*

**(a)** $c$ $\Pr(X = x) = cx^2, x = \{1, 2, 3, 4, 5\}$ is a pmf if the sum over $x$ equals 1. For this to be the case, since $\sum_{x=1}^{5} x^2 = 55$, $c = 1/55 = 0.01818$.

**(b) E and Var**

$$\begin{aligned} \text{E}(X) &= \sum_{x=1}^{5} x \frac{1}{55} x^2 = 4.0909 \\ \text{Var}(X) &= \text{E}(X^2) - (\text{E}(X))^2 = 17.8 - 4.0909^2 = 1.0645 \\ \text{E}(X^2) &= \sum_{x=1}^{5} x^2 \frac{1}{55} x^2 = 17.8 \end{aligned}$$

**Exercise 6** *pdf: c, first two moments, and probability.*

**(a)** $c$ $f_X(x) = cx^2, 1 \leq x \leq 5$ is a pdf if the integral over $x$ equals 1. For this to be the case, since $\int_1^5 x^2 \, \mathrm{d}x = 41.333$, $c = 1/41.333 = 0.024194$.

**(b) E and Var**

$$\begin{aligned} \mathrm{E}(X) &= \int_1^5 x(0.024194)x^2 = 3.7742 \\ \mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 15.116 - 3.7742^2 = 0.87159 \\ \mathrm{E}(X^2) &= \int_1^5 x^2(0.024194)x^2 = 15.116 \end{aligned}$$

**(c) Pr**$(X > 2.0)$

$$\begin{aligned} \Pr(X > 2.0) &= \int_2^5 (0.024194)x^2 \\ &= 0.94355 \end{aligned}$$

**Exercise 7** *cdf.*

$$F_X(x) = \int_0^x \frac{2}{9} u \, \mathrm{d}u = \frac{x^2}{9}$$

Therefore, $F_X(1) = \frac{1^2}{9} = \frac{1}{9}$.

**Exercise 8** *Square transformation of exponential distribution.*

$$f_X(x) = \frac{1}{\tau}e^{-x/\tau}, x \geq 0$$

To perform the transformation for $Y = X^2$, first solve for $X = Y^{\frac{1}{2}}$, then take the derivative $\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{2}y^{-\frac{1}{2}}$ which will surve as the Jacobian for a one-dimensional distribution. Now plug in the expression for $Y$ where $X$ appears in the original pdf, and multiply by the absolute value of that derivative. Lastly, determine the domain for $y$.

$$\begin{aligned} f_Y(y) &= \frac{1}{\tau}e^{-y^{\frac{1}{2}}/\tau}\left|\frac{1}{2}y^{-\frac{1}{2}}\right|, y \geq 0 \\ &= \frac{y^{-\frac{1}{2}}}{2\tau}e^{-y^{\frac{1}{2}}/\tau}, y \geq 0 \end{aligned}$$

**Exercise 9** *Bomb hits.*
Observed frequencies:

| hits | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| freq | 229 | 211 | 93 | 35 | 7 | 1 |

**(a) E** There are a total of 535 bomb hits. To calculate the expected number of hits per area, take the dot product between the number of hits [0..5] with the empirical probabilities of the observed number of hits. Let $K$ be the number of hits for any given area.

$$\mathrm{E}(K) = \sum_{k=0}^{5} k\Pr(k) = \frac{1}{576}(229(0) + 211(1) + 93(2) + 35(3) + 7(4) + 1(5)) = 0.92882$$

**(b) Expected number of hits under "no target ability"** Let the number of hits $K$ follow a Poisson distribution with parameter $\lambda = \frac{535}{576} = 0.92882$, that is, with 576 areas and 535 bombs, each area is expected to be hit $\lambda$ times. (Note: a similar analysis can be done using $K$ following a Binomial$(n = 535, p = \frac{1}{576})$ distribution. The Poisson approximates this well because of the large $n$ and small $p$.)

The domain of the Poisson is countably infinite, but let's consider up to 5 hits, only. The probabilities for 0 to 5 hits are:

```
0.39502   0.3669   0.17039   0.052755   0.01225   0.0022756
```

And the expected number of areas experiencing each number of hits is:

```
211.34   196.29  91.16    28.224    6.5537   1.2174
```

giving a difference between observed and expected:

```
17.6644   14.7074    1.8398    6.7762    0.4463   -0.2174
```

These expected counts under the "no target ability" hypothesis are nearly the same to the observed counts, suggesting that the Germans in fact did not have targetting ability.

To see whether the observed hit frequencies are consistent with the model for "no target ability" we use a chi-square statistics, where obs=observed frequency, exp=expected frequency, and the sum is taken over the 6 cells of the table from 0 to 5 hits.

$$\begin{aligned} X^2 &= \sum_i \frac{(\text{obs} - \text{exp})^2}{\text{exp}} \\ &= 4.3117 \end{aligned}$$

Comparing $X^2 = 4.3117$ to a $\chi^2_{6-1}$ distribution gives a p-value = 0.5055, which is large indicating that the model for "no target ability" is consistent with the observed data.

**Exercise 10** *Difference of dice.*

The probability of each possible outcome is given in the table below, each occuring with probability $\frac{1}{36}$. Therefore, we simply need to count the number of times the random variable (Red−Green) occurs in the table, multiply that by the common probability of $\frac{1}{36}$ and tabulate it.

| | Green | | | | | |
|---|---|---|---|---|---|---|
| Red | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0 | -1 | -2 | -3 | -4 | -5 |
| 2 | 1 | 0 | -1 | -2 | -3 | -4 |
| 3 | 2 | 1 | 0 | -1 | -2 | -3 |
| 4 | 3 | 2 | 1 | 0 | -1 | -2 |
| 5 | 4 | 3 | 2 | 1 | 0 | -1 |
| 6 | 5 | 4 | 3 | 2 | 1 | 0 |

Therefore, the pmf is

$$p_W(w) = \begin{cases} 1/36, & -5 \\ 2/36, & -4 \\ 3/36, & -3 \\ 4/36, & -2 \\ 5/36, & -1 \\ 6/36, & 0 \\ 5/36, & 1 \\ 4/36, & 2 \\ 3/36, & 3 \\ 2/36, & 4 \\ 1/36, & 5 \end{cases}$$

**Exercise 11** *2D Histogram.*

The plots in Figure 1 on page 8 presents the results of the 2D Histograms of Red/Green color intensities with varying numbers of bins. These plots come from output using `hist2.m`, given in the Appendix.

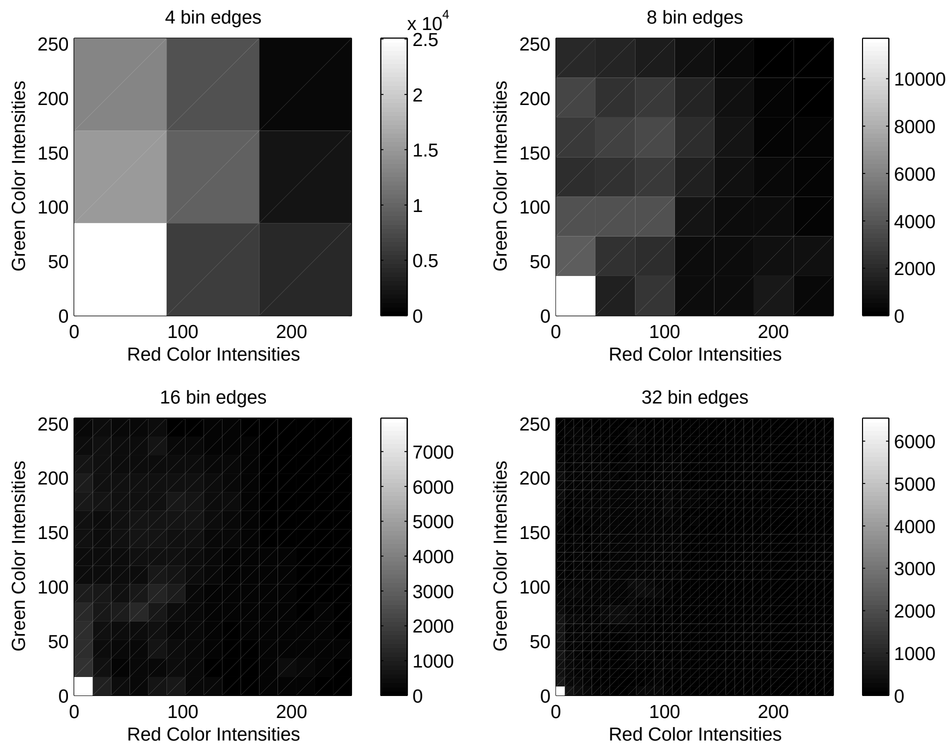


Figure 1: Exercise 11, 2D Histograms of Red/Green color intensities with varying numbers of bins.

**Exercise 12** *2D to 1D Histogram.*

Firstly, the course website program `histogram.m` does not calculate a histogram in the general sense, since it does not allow binning, but instead calculates frequencies of individual values.

To calculate the 1D marginal histograms from a 2D histogram, we need only to sum across rows and columns of the 2D histogram frequencies. This is given in `hist2to1.m`, in the Appendix. Comparing the results to the output of course website program `histogram.m`, I subtract the bin counts and find that they are all 0, showing that the results are identical when using 256 bins. However, my programs have the flexibility of adjustable bin sizes, while the program from the course website only allows bins of size 1.

The plots in Figure 2 on page 9 presents the results of the marginal 1D Histogram of Red/Green color intensities with varying numbers of bins. These plots come from output using `hist2.m` as input to `hist2to1.m`, given in the Appendix.

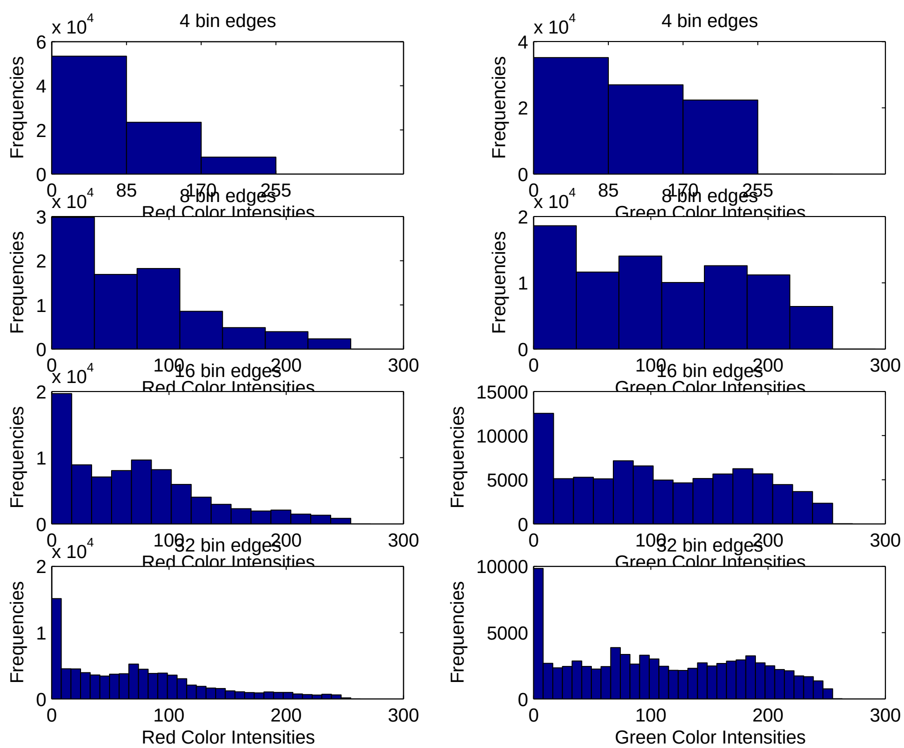


Figure 2: Exercise 12, Marginal 1D Histograms of Red/Green color intensities with varying numbers of bins.

The plots in Figure 3 on page 10 presents the results of the marginal 1D Histogram of Red/Green color intensities with the maximum number of bins, as course website program `histogram.m` would produce.

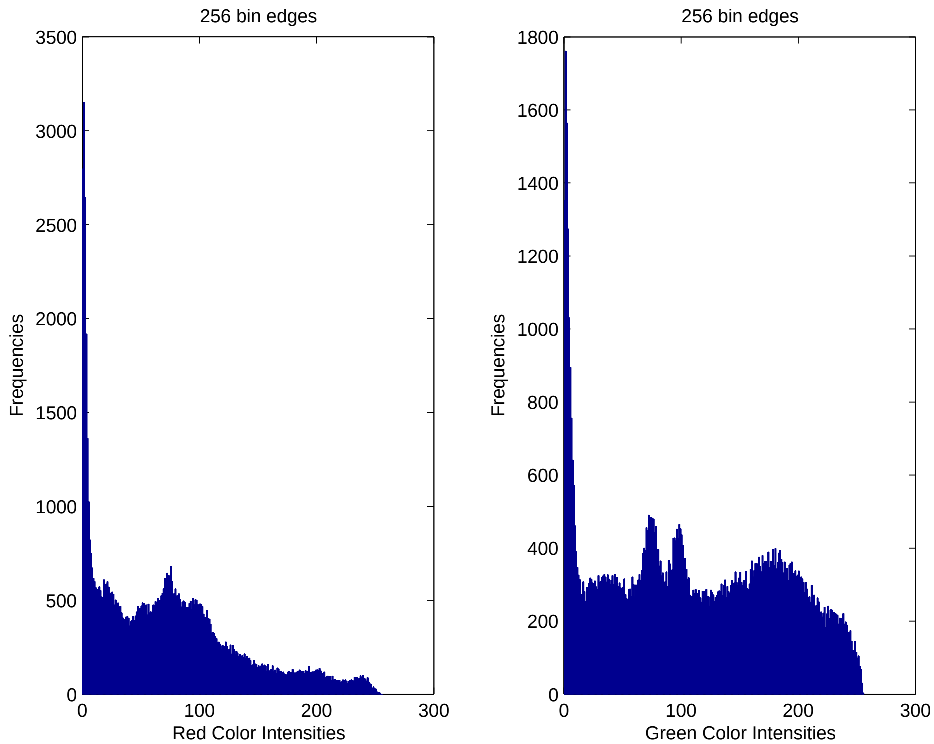


Figure 3: Exercise 11, 2D Histogram of Red/Green color intensities with varying numbers of bins.

# Appendix

## Matlab code used for the exercises

```
%%% Exercise 1  ================================================================
format rat;
x=[5/12 1/18 5/72; 1/36 1/36 5/36; 1/12 1/6 1/72]
x_ind=sum(x,2)*sum(x,1)
x-x_ind

%%% Exercise 2  ================================================================
x=[0 1/9 1/3; 1/18 5/36 1/12; 1/36 1/4 0]
x_x=sum(x,2)
x_y=sum(x,1)
x_xgy=x./repmat(x_y,3,1)
x_ygx=x./repmat(x_x,1,3)

%%% Exercise 3  ================================================================

%%% Exercise 4  ================================================================
p=binopdf([5 6 7 8 9 10],10,0.27)
sum(p)

%%% Exercise 5  ================================================================
sum([1:5].^2)
sum((1/55)*[1:5].^3)
sum((1/55)*[1:5].^4)
sum((1/55)*[1:5].^4) - (sum((1/55)*[1:5].^3))^2

%%% Exercise 6  ================================================================
% a
(5^3-1^3)/3
c=1/((5^3-1^3)/3)
% b
c*((5^4-1^4)/4)
c*((5^5-1^5)/5)
c*((5^5-1^5)/5) - (c*((5^4-1^4)/4))^2
% c
c*((5^3-2^3)/3)

%%% Exercise 7  ================================================================

%%% Exercise 8  ================================================================

%%% Exercise 9  ================================================================
obs=[229 211 93 35 7 1];
[0:5;obs] % print observed freqs
n=[0:5]*[229 211 93 35 7 1]' % number of hits
[0:5]*([229 211 93 35 7 1]'/n) % average number of hits

535/576
poisspdf([0:5],535/576)
ex=535*poisspdf([0:5],535/576)

obs-ex
chi2=sum((obs-ex).^2./ex)  % chi2 statistic
p=1-chi2cdf(chi2,6-1)  % p-value

%%% Exercise 10 ================================================================
format rat
(repmat([1:6]',1,6)-repmat([1:6],6,1)) % row - col

%%% Exercise 11 ================================================================
```

Listing 1: Matlab routine `hist2.m` used to generate 2D joint histogram.

```
function histmat = hist2(x, y, xedges, yedges)
% function histmat  = hist2(x, y, xedges, yedges)
%
% Construct a matrix of 2 dimensional histogram frequencies of (x,y) pairs
% falling within the bins of the grid defined by xedges and yedges.
% The edges are vectors with monotonically non-decreasing values.
%  Construct the edges using linspace() or logspace(), eg:
%     xedges = linspace(-1,1,64);
%     xedges = logspace(0,log10(3),64)-2;
%
% Example of use:
%  Generate simulated data
%   n = 50;
%   x1 = randn(n,1)-1; x2 = randn(n,1)+1;
%   y1 = randn(n,1)+1; y2 = randn(n,1)-1;
%   x= [x1;x2]; y = [y1;y2];
%  Use function
```

```
%   xedges = linspace(-1,1,5); yedges = linspace(-1,1,5);
%   histmat = hist2(x, y, xedges, yedges);
%   figure; pcolor(xedges, yedges, histmat'); colorbar; axis square tight;
%
%  Erik Barry Erhardt 8/23/2006

% error checking
if nargin~=4;               error('Four input arguments required!'); return; end;
if any(size(x)~=size(y));   error('The x and y dimensions must be same!'); return; end;

% Use histc to count the number of observations in the marginal directions
[xn, xbin]=histc(x, xedges);        % x vector,    xbin and ybin are vectors indicating
[yn, ybin]=histc(y, yedges);        % y vector,       which bin the observation belongs

% Count the number of bins in the x and y dimensions
xnbin=length(xedges);
ynbin=length(yedges);

% Assigns each observation to a cell within the xnbin*ynbin histogram matrix (as a vector)
if xnbin>=ynbin;
   xy=ybin*xnbin+xbin;
else;
   xy=xbin*ynbin+ybin;
end;

% Final count
xyuni=unique(xy);                   % gets the cell indicies where data exists
hstres=histc(xy,xyuni);             % counts the number of observations in those cells
histmat=zeros(xnbin,ynbin);         % creates a matrix with the correct number of rows/columns
histmat(xyuni-xnbin)=hstres;        % assigns the counts to those indicies

% EOF %
```

```
% Read in ppm image data
%  Create a matrix (width*height,3), with the three columns being (Red Green Blue)

fn='queen_butterfly_fish.ppm';
fid=fopen(fn,'r');
   junk=fgetl(fid);                  % read off first two lines
   junk=fgetl(fid);                  % which don't contain important information
   im=fscanf(fid,'%3d %3d %3d');
fclose(fid);

width=im(1); height=im(2); depth=im(3);          % image sizes
X=reshape(im(4:end),length(im(4:end))/3,3);     % create a three-column matrix
R=X(:,1); G=X(:,2); B=X(:,3);                   % Red, Green, and Blue

% Perform 2D histogram on Red and Green components

figure(10)
for i=2:5;
   nbins=2^i;
   xedges = linspace(0, 255, nbins);
   yedges = linspace(0, 255, nbins);
   x=R;y=G;
   histmat = hist2(x, y, xedges, yedges);
   subplot(2,2,i-1);
      pcolor(xedges, yedges, histmat');
      colorbar; axis square tight; colormap(gray); shading flat;
   title([num2str(2^i),' bin edges']);
   xlabel('Red Color Intensities');
   ylabel('Green Color Intensities');
end;
plot_name = strcat('gpcshw01-11.eps'); % plot name
print(gcf, '-depsc2', plot_name);   % print plot

   %% Other viewing techniques:
      %% image
      % image(histmat')
      % imagesc(histmat')
      %% contour plots
      % contour(xedges, yedges,histmat')
      % contour3(xedges, yedges,histmat')
      %% surface plot
      % surf(xedges, yedges,histmat')
      % surfc(xedges, yedges,histmat')  % rotate to see contours
      % surfl(xedges, yedges,histmat')  % shading from a light source
      %% mesh plots
      % mesh(xedges, yedges,histmat')
      % meshc(xedges, yedges,histmat')  % rotate to see contours
      % waterfall(xedges, yedges,histmat')

%%% Exercise 12 ================================================================
```

Listing 2: Matlab routine `hist2to1.m` to compute marginals from 2D histogram data.

```
function [xmarg, ymarg] = hist2to1(histmat)
% function [xmarg, ymarg] = hist2to1(histmat)
%
% Construct the marginal histograms from the 2D histogram from hist2.m
%
% Example of use:
%  Use function
%   histmat = hist2(x, y, xedges, yedges);
%   [xmarg, ymarg] = hist2to1(histmat);
%
%  Erik Barry Erhardt 8/23/2006

% error checking
if nargin~=1;             error('Only one input argument!'); return; end;

xmarg=sum(histmat,2);
ymarg=sum(histmat,1);

% EOF %
```

Listing 3: Matlab routine `histogram.m` from course website.

```
function h=histogram(f,N)

% histogram -- computes a grey level histogram.
% version 1
% Lance Williams
% Dept. of Computer Science
% Univ. of New Mexico
%

for i=0:N-1
  h(i+1) = sum(sum(f==i));
end
```

```
% Perform marginal 1D histograms on Red and Green components for different bin sizes
figure(11)
for i=2:5;
   nbins=2^i;
   xedges = linspace(0, 255, nbins);
   yedges = linspace(0, 255, nbins);
   x=R;y=G;
   histmat = hist2(x, y, xedges, yedges);
   % hist2to1 calculating marginal histograms from joint histogram
   [xmarg, ymarg] = hist2to1(histmat);
   subplot(4,2,2*i-3);
      bar(xedges,xmarg,'histc');
      title([num2str(2^i),' bin edges']);
      xlabel('Red Color Intensities');
      ylabel('Frequencies');
   subplot(4,2,2*i-2);
      bar(yedges,ymarg,'histc');
      title([num2str(2^i),' bin edges']);
      xlabel('Green Color Intensities');
      ylabel('Frequencies');
end;
plot_name = strcat('gpcshw01-12a.eps'); % plot name
print(gcf, '-depsc2', plot_name);   % print plot


% every intensity seperately
figure(12)
nbins=256;
xedges = linspace(0, 255, nbins);
yedges = linspace(0, 255, nbins);
x=R;y=G;
histmat = hist2(x, y, xedges, yedges);
% hist2to1 calculating marginal histograms from joint histogram
[xmarg, ymarg] = hist2to1(histmat);
subplot(1,2,1);
   bar(xedges,xmarg,'histc');
   title('256 bin edges');
   xlabel('Red Color Intensities');
   ylabel('Frequencies');
subplot(1,2,2);
   bar(yedges,ymarg,'histc');
```

```
   title('256 bin edges');
   xlabel('Green Color Intensities');
   ylabel('Frequencies');
plot_name = strcat('gpcshw01-12b.eps'); % plot name
print(gcf, '-depsc2', plot_name);   % print plot



% To compare my 2D to 1D method to course website histogram method,
%  use seperate bins for every possible outcome
nbins=256;
xedges = linspace(0, 255, nbins);
yedges = linspace(0, 255, nbins);
x=R;y=G;
histmat = hist2(x, y, xedges, yedges);

% hist2to1 calculating marginal histograms from joint histogram
[xmarg, ymarg] = hist2to1(histmat);
histogram(histmat)

% histogram program from course website
xmarg2=histogram(R,256); % Red
ymarg2=histogram(G,256); % Green

% Compare hist2to1 to histogram
sum(abs(xmarg'-xmarg2)) % =0, so no difference
sum(abs(ymarg -ymarg2)) % =0, so no difference
```