Week 2: Probability review
Bernoulli, binomial, Poisson, and normal distributions
Solutions

A Binomial distribution. To evaluate the mean and variance of a binomial RV B_n with parameters (n, p), we will rely on the relation between the binomial and the Bernoulli. First, let $\{L_i\}_{i=1,\dots,n}$ be independent Bernoulli RVs with probability of success p. Then, the expected value of a single L_i can be computed as

$$\mathbb{E}[L_i] = 1 \cdot p + 0 \cdot (1-p) = p. \tag{1}$$

Indeed, the expected value is the sum of possible outcomes weighted by their probabilities by definition. Using the fact that B_n is the sum of n independent and identically distributed (i.i.d.) Bernoulli RVs L_i, we can derive its expected value from (1) by linearity:

$$\begin{aligned} B_n = \sum_{i=1}^{n} L_i \Rightarrow \mathbb{E}[B_n] = \mathbb{E}\left[\sum_{i=1}^{n} L_i\right] = \sum_{i=1}^{n} \mathbb{E}[L_i] \quad \text{[from the linearity of expectation]} \\ = \sum_{i=1}^{n} p = np. \end{aligned}$$

We can proceed in a similar way for the variance. First, we compute some second-order moments of Bernoulli RVs. Explicitly,

$$\begin{aligned} \mathbb{E}\left[L_i^2\right] &= 1^2 \cdot p + 0^2 \cdot (1-p) = p, \\ \mathbb{E}[L_i L_j] &= \mathbb{E}[L_i] \cdot \mathbb{E}[L_j] = p \cdot p = p^2, \end{aligned}$$

where we recall that $\{L_i, L_j\}$ are independent for $i \neq j$. Then,

$$\begin{aligned} \operatorname{var}(B_n) = \mathbb{E}\left[(B_n - \mathbb{E}(B_n))^2\right] = \mathbb{E}\left[B_n^2\right] - (\mathbb{E}[B_n])^2 = \mathbb{E}\left[\left(\sum_{i=1}^{n} L_i\right)\left(\sum_{j=1}^{n} L_j\right)\right] - (np)^2 \\ = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[L_i L_j] - n^2p^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p). \end{aligned}$$

A MATLAB script that computes the values of the binomial pmf is presented below.

```
% Takes vector k and scalars n and p and returns a vector of the same size
% as k, whose entries represent the probability of a binomial RV with
% parameters n and p at the points of the elements of k.

function pmf = binomial_pmf(k, n, p)

```

```
pmf = zeros(size(k)); % Initialize output

for i = 1:length(k)
    % Check that the element of k is in the support of the binomial(n,p).
    % Otherwise, pmf = 0.
    if (k(i) >= 0) && (k(i) <= n)
        pmf(i) = nchoosek(n,k(i)) * p^k(i) * (1-p)^(n-k(i));
    end
end

end
```

You could write a similar code replacing **nchoosek** by the function **factorial**, but be aware that it is only accurate for numbers up to 21. This is because MATLAB represents numbers in double precision which have roughly 15 "usable" digits. The function **nchoosek** uses different approximations for large n.

We can use **binomial_pmf.m** to plot the pmfs and cdfs requested in part A. Notice we use of **stem** to show the pmf and **stairs** for the cdf, since these are discrete RV.

```
% Delete all variables and close figures
clear all
close all

n_vector = [6, 10, 20, 50];

for i = 1:length(n_vector)
    n = n_vector(i);
    p = 5/n;            % E[B_n] = 5

    h1 = figure(1);
    subplot(2,2,i);
    stem(0:n, binomial_pmf(0:n, n, p), '.');
    title(['n = ' num2str(n)]);
    xlabel('k');
    ylabel('pmf');
    grid;
    xlim([0,50]);
    ylim([0,0.5]);

    h2 = figure(2);
    subplot(2,2,i);
    stairs(0:n, cumsum(binomial_pmf(0:n, n, p)), 'LineWidth', 2);
    title(['n = ' num2str(n)]);
    xlabel('k');
    ylabel('cdf');
    grid;
    xlim([0,50]);
    ylim([0,1]);
end

%%% Export figure %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
set(h1,'Color','w');
export_fig('-q101', '-pdf', 'HW2_A.pdf', h1);
set(h2,'Color','w');
export_fig('-q101', '-pdf', '-append', 'HW2_A.pdf', h2);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The pmfs and cdfs are show in Figures 1 and 2.

B Binomial and Poisson distributions. The expected value of a Poisson RV P with parameter λ is given by

$$\begin{aligned}\mathbb{E}[P] &= \sum_{k=0}^{\infty} k \cdot \left(\frac{e^{-\lambda}\lambda^k}{k!}\right) = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{d}{d\lambda}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) \\ &= \lambda e^{-\lambda} \frac{d}{d\lambda}\left(e^{\lambda}\right) = \lambda e^{-\lambda} e^{\lambda} = \lambda.\end{aligned}$$

Notice that in the first equality in the second line comes from recognizing that the infinite sum is the Taylor series of the exponential.

The MATLAB code to plot the Poisson distribution is shown below.

```
% Delete all variables and close figures
clear all
close all

lambda = 5;
k = 0:50;

poisson_pmf = exp(-lambda) * (lambda.^k) ./ factorial(k);
% Note that we use the "dot" to perform elementwise operations and evaluate
% the pmf for points in k at once.

figure();
stem(k, poisson_pmf, '.');
title(['Poisson pmf with \lambda = ' num2str(lambda)]);
xlabel('k');
ylabel('pmf');
grid;


%%% Export figure %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
set(gcf,'Color','w');
export_fig('-q101', '-pdf', 'HW2_B1.pdf');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The result can be found in Figure 3. Note the similarity with Figures 1 for large n.

To calculate the MSE, we first check a few values of the Poisson pmf to find the cut-off point for the exercise. Simple trial and error shows that we need only considers values for $k \in [3, 9]$. Hence, we can use the following code to evaluate and plot the MSE.

```
% Delete all variables and close figures
clear all
close all

lambda = 5;
k = 3:9;

poisson_pdf = exp(-lambda)*(lambda.^k) ./ factorial(k);

n_vector=[6 10 20 50];
mse = zeros(length(n_vector), 1);

for i=1:length(n_vector)
    n = n_vector(i);
    mse(i) = sum( (binomial_pmf(k, n, lambda/n) - poisson_pdf).^2 .* poisson_pdf );
end

figure();
```

```
plot(n_vector, mse, 'x', 'LineWidth', 2);
title('MSE between Poisson and binomial pmfs')
xlabel('n');
ylabel('MSE');
grid;


disp(mse);


%%% Export figure %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
set(gcf,'Color','w');
export_fig('-q101', '-pdf', 'HW2_B2.pdf');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The resulting plot is shown in Figure 4. Values are reported in Table 1.

Table 1: MSE between the pmf of a Poisson with $\lambda = 5$ and binomials with parameters $(n, \lambda/n)$ for $n = 6, 10, 20, 50$.

n	MSE
6	0.01684
10	0.00168
20	0.00025
50	0.00003

The table above shows that the MSE between the binomial and Poisson distributions rapidly decreases to zero as n increases, indicating that the distributions become increasingly similar for large n.

C Binomial and Poisson distributions again. We start by writing out the pmf of B_n. Explicitly,

$$\mathbb{P}[B_n = k] = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Then, notice that we can rearrange the first two terms to read

$$\begin{aligned}\mathbb{P}[B_n = k] &= \frac{\lambda^k}{k!} \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n}{n} \frac{(n-1)}{n} \frac{(n-2)}{n} \cdots \frac{(n-k+1)}{n} \left(1 - \frac{\lambda}{n}\right)^{n-k}\end{aligned}$$

Finally, we can take the limit as $n \to \infty$. Recall that the limit of the product is the product of the limit as long as all limits exist, which is the case here:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}[B_n = k] &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n}{n} \frac{(n-1)}{n} \frac{(n-2)}{n} \cdots \frac{(n-k+1)}{n} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} (1 \times 1 \times \cdots \times 1) \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k}.\end{aligned}$$

The last limit is simply 1^{-k}. Also, by definition, we have

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Thus,

$$\lim_{n\to\infty}\mathbb{P}\left[B_n = k\right] = \frac{\lambda^k}{k!}e^{-\lambda} = \mathbb{P}\left[P = k\right].$$

D Binomial and normal distributions. If Z_n is a normal RV with zero mean and unit variance (also known as a *standard normal*), then from

$$Z_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}},$$

it holds that $Y_n = \sum_{i=1}^n Y_i$ is also normally distributed but with mean np and variance $n\sigma^2$, where $\sigma^2 = p(1-p)$ is the variance of a single Bernoulli Y_i. The MATLAB code to display this approximation is given below.

```
% Delete all variables and close figures
clear all
close all

n_vector = [10 20 50];
p = 0.5;

for i = 1:length(n_vector)
    n = n_vector(i);

    Yn_mean = n*p;
    Yn_var = n*p*(1-p);
    k = 0:n;

    figure(1);
    subplot(3,1,i);
    stairs(k, cumsum(binomial_pmf(k, n, p)), 'LineWidth', 2);
    hold on
    stairs(k, normcdf(k, Yn_mean, sqrt(Yn_var)), 'LineWidth', 2);
    title(['n=',num2str(n)]);
    xlabel('k');
    ylabel('cdf');
    legend('Binomial', 'Normal');
    grid;
end



%%% Export figure %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
set(gcf,'Color','w');
export_fig('-q101', '-pdf', 'HW2_D.pdf');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The resulting plot can be found in Figure 5.

E Normal and Poisson approximations. I'll provide you with a hint for this part: The Poisson limit theorem is about counting a large number of increasingly *improbable* events. In particular, note that for the distribution of a sum of i.i.d. Bernoulli RVs (i.e., a binomial RV) to converge to a Poison distribution with mean λ, the probability of success of each Bernoulli trial

must be $p = \lambda/n$, which goes to zero as $n \to \infty$. On the other hand, p is fixed for the CLT. Hence, the CLT and the Poisson limit theorem are addressing fundamentally different limits. I leave more contemplation on this matter to you!

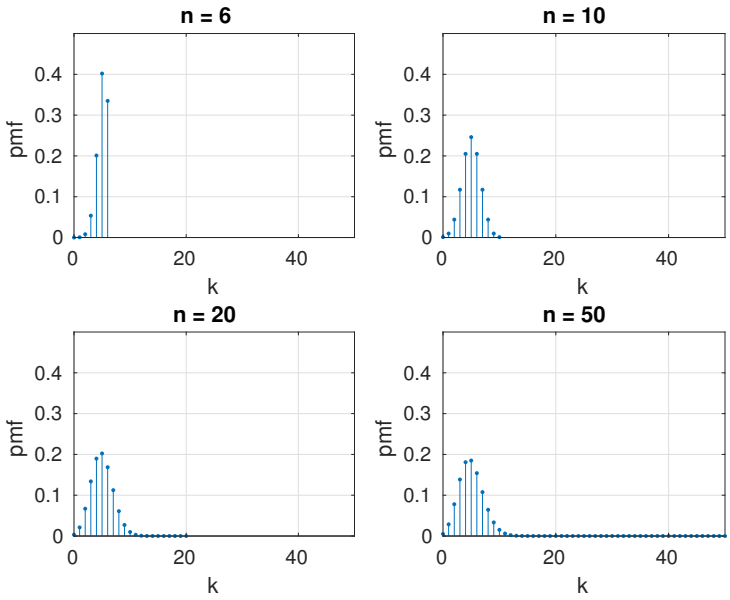

Figure 1: Binomial pmf for $n = 6, 10, 20, 50$ and $p = 5/n$ (part A).

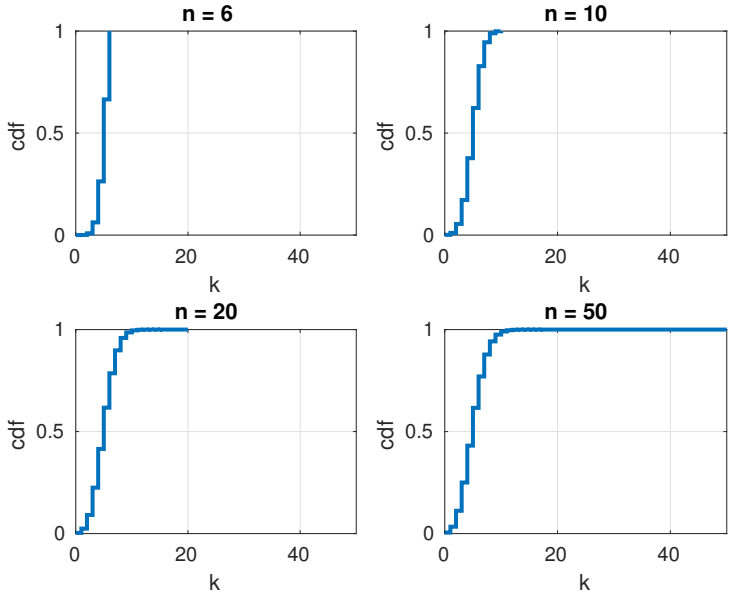

Figure 2: Binomial cdf for $n = 6, 10, 20, 50$ and $p = 5/n$ (part A).

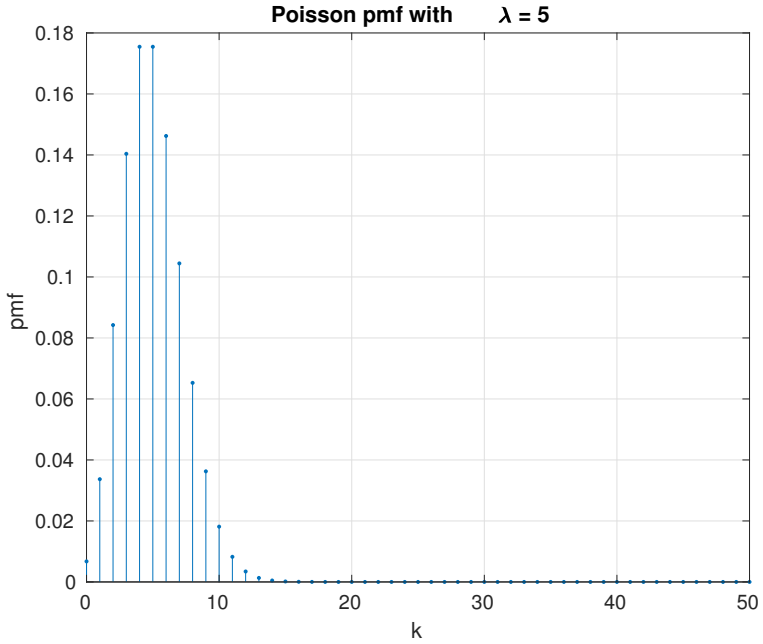

Figure 3: The pmf of a Poisson RV with $\lambda = 5$ (part B). Note that the support of the Poisson distribution is the whole non-negative integer line. However, we display only the first 50 points.

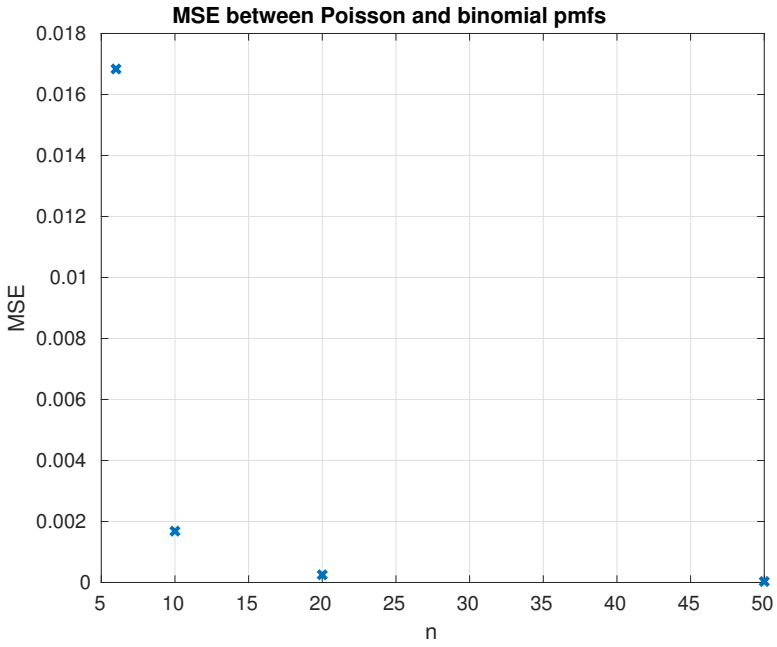

Figure 4: MSE between the pmf of a Poisson with $\lambda = 5$ and binomials with parameters $(n, \lambda/n)$ for $n = 6, 10, 20, 50$ (part B).

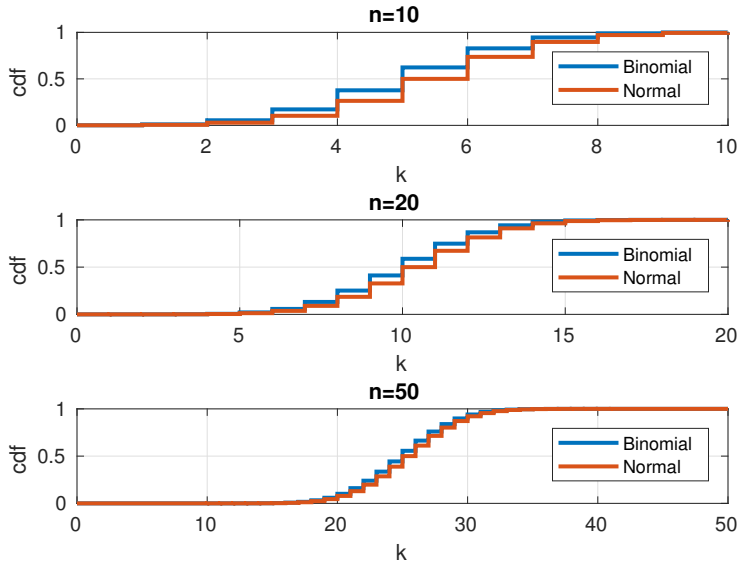

Figure 5: Cumulative distribution function of the binomial RV Y_n for $n = 10, 20, 30$ and its approximation by a normal cdf (part D).