Sample Problems
Lecture 2: Probability Distributions

Problems prepared by: Prof. F. Jalayer and Dr. H. Ebrahimian

1. The rainfall problem

Suppose that there is a probability equal to 0.20 for having rain in a day within the winter, and assume that having rain on consecutive days is independent.

- **What is the probability of having rainy day once in a given week?**

$\mathbf{R} \equiv$ it is a rainy day (i.e. having at least one significant rainfall)

$\overline{\mathbf{R}} \equiv$ it is not a rainy day

The probability that the statement **R** is TRUE is equal to:

$$P\left(\mathbf{R}\middle|\mathbf{I}\right) = \pi = 0.20$$

and consequently

$$P\left(\overline{\mathbf{R}}\middle|\mathbf{I}\right) = 1 - \pi = 1 - 0.20 = 0.80$$

$$P(r = 1 \mid n = 7, \pi = 0.2) = \binom{n}{r}\pi^r\left(1-\pi\right)^{n-r} = \binom{7}{1}0.20^1\left(1-0.20\right)^6 \cong 0.37$$

- **What is the probability of having at least one rainy day in a week?**

$$P(r >= 1 \mid n = 7, \pi = 0.2) = 1 - P\left(r = 0\middle|n = 7, \pi = 0.2\right) = 1 - \binom{n}{r}\pi^r\left(1-\pi\right)^{n-r}$$
$$= 1 - \binom{7}{0}0.20^0\left(1-0.20\right)^7 \cong 0.79$$

- **What is the probability of having a rainy day at least twice a week?**

$$P(r >= 2 \mid n = 7, \pi = 0.2) = 1 - P\left(r = 0 \middle| n = 7, \pi = 0.2\right) - P\left(r = 1 \middle| n = 7, \pi = 0.2\right)$$
$$\cong 1 - 0.21 - 0.37 \cong 0.42$$

- **What is the probability of having rain tomorrow, given that today is rainy?**

$$P(\text{tomorrow will be rainy} \mid \text{today is rainy}) = P(\text{tomorrow will be rainy}) = \pi = 0.20$$

- **What is probability of having all sunny days during the February?**

$$P(r = 0 \mid n = 29, \pi = 0.2) = 0.0015$$

2. The coin problem

Consider that a coin is tossed many times. The coin has two faces: A and B. Let the probability of having A to be equal to 0.55.

- **What is the probability of having face B three times consecutively?**

This problem can be interpreted as having three successes (r=3) (face B) out of three trials (n=3). The probability of having B is equal to 1-0.55=0.45. Therefore, by using the Binomial distribution to calculate the probability of 3 "successes" in a row out of three trials:

$$P\left(r=3\middle|n=3,\pi=0.45\right)=\binom{3}{3}0.45^{3}\left(1-0.45\right)^{0}=0.091$$

- **What is the probability of having at least two times the face B in 10 trials?**

The CCDF of the binomial distribution:

$$P\left(r\geq 2\middle|n=10,\pi=0.45\right)=1-\binom{10}{0}0.45^{0}\left(1-0.45\right)^{10}-\binom{10}{1}0.45^{1}\left(1-0.45\right)^{9}=0.9767$$

- **Considering the two following hypothesis:**

  H_1: P(face A) = 0.5 H_2: P(face A) = 0.60

  Suppose that Face A comes out of three trials. What is the weight that can be assigned to the hypothesis H_2 with respect to the hypothesis H_1?

The weight of the hypothesis H_2 with respect to H_1 is given by the ratio between the conditional probability of having H_2 given the new knowledge level D to the conditional probability H_1, (where the statement D = having three times face A out of three trials):

$$\frac{P\left(H_2\mid D\right)}{P\left(H_1\mid D\right)}=\frac{P\left(D\mid H_2\right)\cdot P\left(H_2\right)}{P\left(D\mid H_1\right)\cdot P\left(H_1\right)}\approx\frac{P\left(D\mid H_2\right)}{P\left(D\mid H_1\right)}$$

Both hypotheses have the same plausibility a priori; hence, the two terms $P(\mathrm{H}_2)$ and $P(\mathrm{H}_1)$ are equal and cancel out as demonstrated above. As a result, the two terms that remains, $P(\mathrm{D}|\mathrm{H}_2)$ and $P(\mathrm{D}|\mathrm{H}_1)$, are the probability of having three success out of three trials with the binomial distribution, considering hypotheses H_2 and H_1 respectively:

$$\frac{P(H_2 \mid D)}{P(H_1 \mid D)} = \frac{P(D \mid H_2)}{P(D \mid H_1)} = \frac{\binom{3}{3} \cdot (0.6)^3 \cdot (1-0.6)^0}{\binom{3}{3} \cdot (0.5)^3 \cdot (1-0.5)^0} = 1.728$$

Therefore, the observed data seem to tell us that the odds of having hypothesis H_2 be TRUE (i.e. $P(\mathrm{A})=0.6$) is 1.72 times the odds of having hypothesis H_1 be TRUE.

3. Compressive test of cylindrical concrete specimens

Suppose that we have tested the compressive strength of n cylindrical concrete specimens. Let the Hypothesis H being defined as the probability that the cylindrical strength of a specimen, f_c, is greater than a pre-defined threshold of 20 MPa is equal to π.

Consider that our test results indicate (statement D) that we have had r specimens that satisfy the hypothesis H (i.e., their resistance is greater than 20Mpa).

- **What is the probability of having *r* specimens with resistance greater than 20 MPa given that we know the Hypothesis H (hint: calculate *P*(*D*|*H*,*I*))?**

The statements can be defined as:

D: having r specimens with resistance greater than 20Mpa out of n tests

H: the probability that f_c>20Mpa is equal to π

I: other background information

$$P(D \mid H, I) = P(r \mid n, \pi) = \binom{n}{r} \pi^r (1-\pi)^{n-r}$$

- **What is the probability that can be assigned to Hypothesis H given the test results (i.e., statement D)? (use Bayes Theorem)**

$$P(H \mid D, I) = \frac{P(D \mid H, I) P(H \mid I)}{P(D \mid I)} = c^{-1} P(D \mid H, I) P(H \mid I)$$

$P(H \mid I) = P(\pi \mid I) = 1$ Uniform distribution (all values of π have the same likelihood)

$$\Rightarrow P(H \mid D, I) = \frac{\binom{n}{r} \pi^r (1-\pi)^{n-r}}{\int_0^1 \binom{n}{r} \pi^r (1-\pi)^{n-r} \, \mathrm{d}\pi} = \frac{\pi^r (1-\pi)^{n-r}}{\int_0^1 \pi^r (1-\pi)^{n-r} \, \mathrm{d}\pi}$$

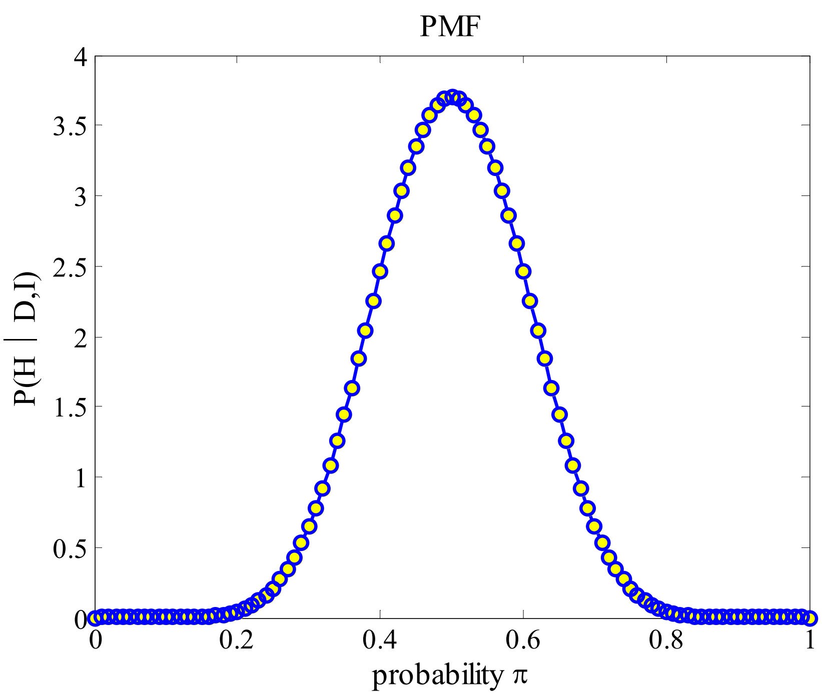
The integral in the denominator is called in mathematics *the Euler integral of the first kind*, which is also known as the *Beta function*. It is a special function defined as:

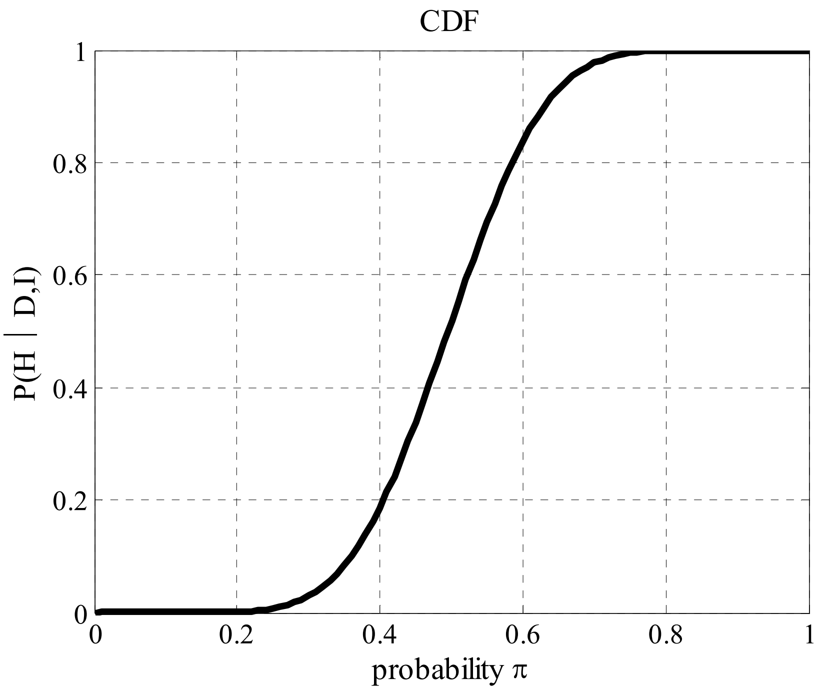
$$\int_0^1 \pi^r \left(1-\pi\right)^{n-r} \mathrm{d}\pi = \frac{r!\left(n-r\right)!}{\left(n+1\right)!}$$

Thus,

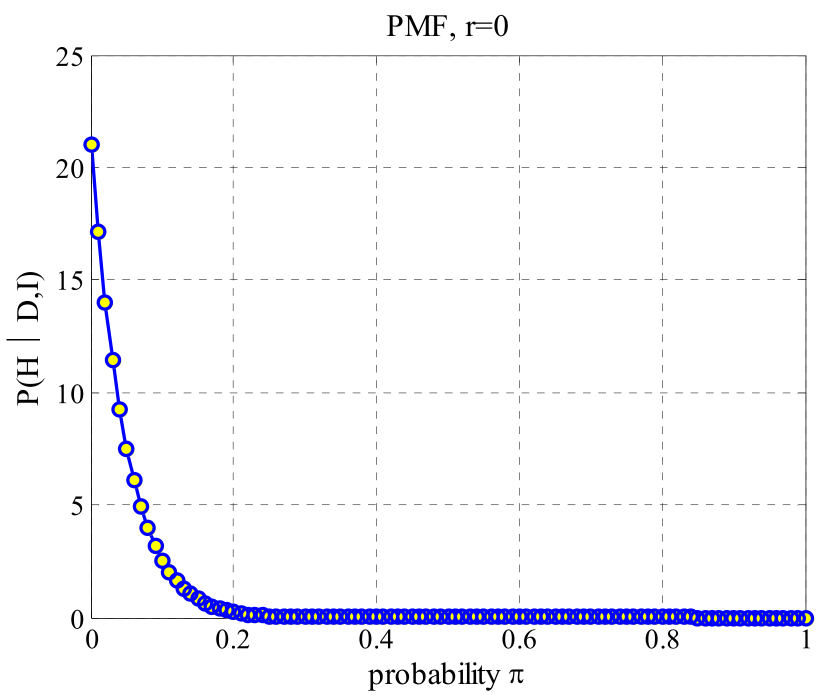
$$P\left(H \mid D, I\right) = \frac{\left(n+1\right)!}{r!\left(n-r\right)!}\pi^r \left(1-\pi\right)^{n-r}$$

- **Plot the PMF and CDF associated with the probability that Hypothesis H is TRUE given the test results for the domain of probabilities π varying from zero to one, considering r=10 and n=20?**





- **Plot the PMF associated with the probability of Hypothesis H given the knowledge level for the boundary values of *r*=0?**

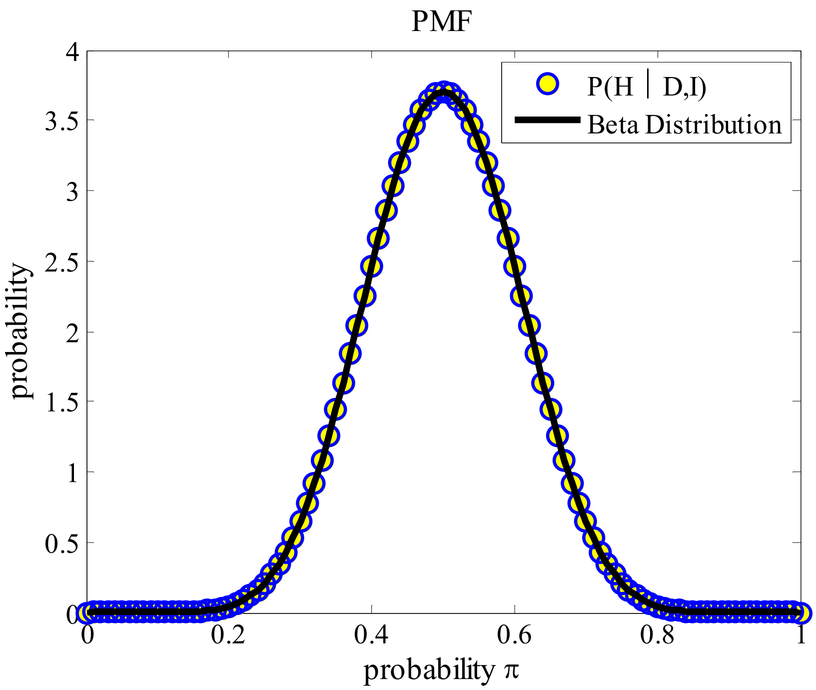


Note: It can be shown that the posterior probability that you have obtained in the second part is actually a "Complete Beta Distribution".

The probability density function (PDF) of the Beta distribution, for $0 \le x \le 1$, and shape parameters α, $\beta > 0$, is actually derived by a power function of the variable x and of its reflection $(1-x)$ as follows:

$$
\begin{aligned}
f\left(x \middle| \alpha,\beta\right) &= \text{constant}\cdot x^{\alpha-1}\cdot\left(1-x\right)^{\beta-1} \\
&= \frac{x^{\alpha-1}\left(1-x\right)^{\beta-1}}{\int_0^1 y^{\alpha-1}\left(1-y\right)^{\beta-1}\,\mathrm{d}y} = \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}x^{\alpha-1}\left(1-x\right)^{\beta-1} = \frac{\left(\alpha+\beta-1\right)!}{\left(\alpha-1\right)!\left(\beta-1\right)!}x^{\alpha-1}\left(1-x\right)^{\beta-1} \\
&= \frac{1}{\mathrm{B}\left(\alpha,\beta\right)}x^{\alpha-1}\left(1-x\right)^{\beta-1}
\end{aligned}
$$

where Γ is called the *Gamma* function and B is the *Beta* function, which is a normalization constant to ensure that the total probability integrates to 1. With a similar analogy, if we assign $\alpha = r+1,\ \beta = n-r+1$, the Beta distribution is equal to the posterior probability $P(D|H,I)$ estimated in the second part of this problem. $P(H|D,\mathrm{I})$ from the third part (with r=10, n=20) is compared with the generalized Beta distribution from the above formula:


PMF
P(H | D,I)
Beta Distribution
probability
probability π
0
0.5
1
1.5
2
2.5
3
3.5
4
0.2
0.4
0.6
0.8

4. Construction defects

The recent European codes provide a level of conservatism in the assessment of existing buildings, by the application of the confidence factors (CF) to mean material property estimates. These confidence factors are to be determined as a function of the knowledge levels (KL). The knowledge levels are determined based on the amount of tests and inspections performed on the building. The following table illustrates the three KL's, namely, limited, extended and comprehensive:

KL	Inspections of reinforcement details (% structural elements)
Limited	20
Extended	50
Comprehensive	80

One of the main sources of uncertainty is due to the uncertainty in structural construction details. In particular, the structural construction details can include stirrup spacing, concrete cover, anchorage, and splice length; these are also known as the *defects* when they present deviations from the original configurations leading to undesirable effects.

- **In a building with N=20 columns, let the probability of having a defect in a given column be equal to p. Imagine that we have done enough in-situ tests to satisfy the limited knowledge level (i.e., 20% of the columns) and have observed no defect. What is the probability distribution for p? (Assume that having defect in different columns are independent events).**

The statements can be defined as:

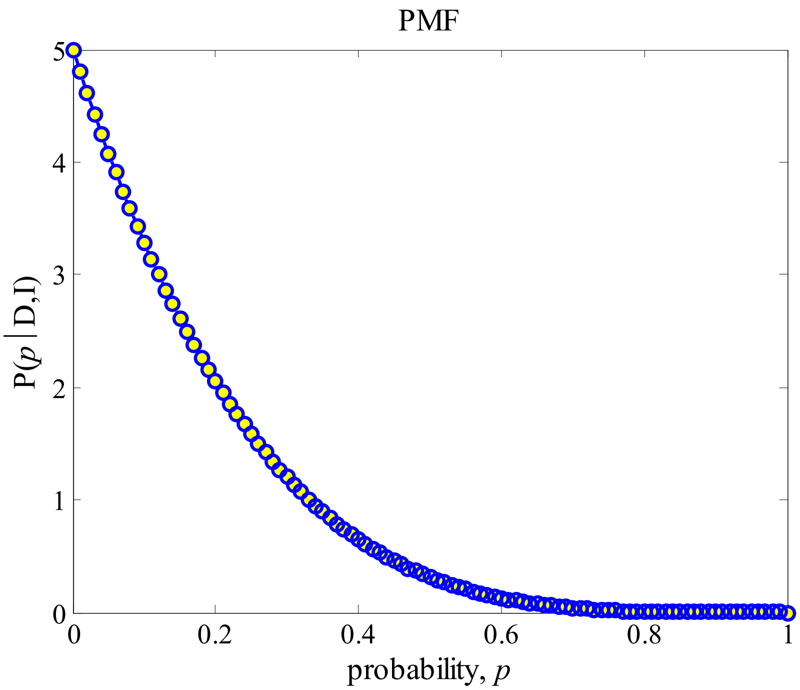
D: out of n=0.20N columns tested, n_d=0 of them demonstrate a defect

H: the probability of having a defect in a given column be equal to p

I: the building has N columns and the in-situ tests to satisfy the limited knowledge level

$$P(D|H,I) = P(n_d|n,p) = \binom{n}{n_d} p^{n_d}(1-p)^{n-n_d}$$

$$P(H|D,I) = \frac{P(D|H,I)P(H|I)}{P(D|I)} = c^{-1}\binom{n}{n_d} p^{n_d}(1-p)^{n-n_d} = \frac{(n+1)!}{n_d!(n-n_d)!} p^{n_d}(1-p)^{n-n_d}$$

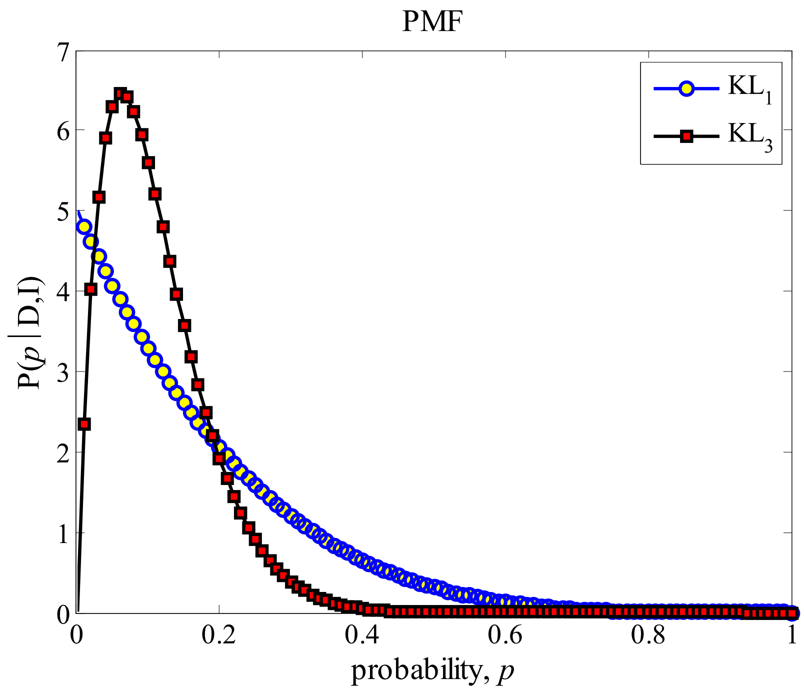


- **Imagine that we have done enough in-situ tests to satisfy the comprehensive knowledge level (i.e., 80% of the columns) and have observed one defect. What is the probability distribution for *p*? (Assume that having defect in different columns are independent events).**

D: out of n=0.80N columns tested, n_d=1 of them demonstrate a defect

H: the probability of having a defect in a given column be equal to p

I: the building has N columns and the in-situ tests to satisfy the limited knowledge level



- **Imagine that you know that you have a total of *M* columns with defect in the whole building and that we have done enough in-situ tests to satisfy the comprehensive knowledge level (i.e., 80% of the columns). What is the probability that the first column inspected is defected?**

$D_i \equiv$ The ith column inspected is with defect

$$P(D_1) = \frac{M}{N}$$

- **What is the probability that the first two columns inspected are both with a defect?**

$$P(D_1 D_2) = P(D_1) P(D_2 \mid D_1) = \frac{M}{N} \times \frac{M-1}{N-1} \text{ (Sampling without replacement)}$$

- **What is the probability that all of the first *r* columns inspected are with defects?**

$$P(D_1 D_2 \cdots D_r) = \frac{M}{N} \times \frac{M-1}{N-1} \times \cdots \times \frac{M-r+1}{N-r+1} = \frac{\dfrac{M!}{(M-r)!}}{\dfrac{N!}{(N-r)!}} = \frac{M!(N-r)!}{(M-r)!N!}$$

- **What is the probability of that all of the first *w* columns inspected are without defects? (hint: it can be resolved by analogy with the previous problem)**

$ND_i \equiv$ the ith column inspected is without defect

Interchange M with N-M, and r with w, in the previous example:

$$P(ND_1 ND_2 \cdots ND_w) = \frac{(N-M)!\,(N-w)!}{(N-M-w)!\,N!}$$

- **What is the probability that only the first *r* columns inspected in a sample of *n*=0.8*N* columns are with defects?**

The problem can be interpreted as the probability of obtaining first r columns having defect followed by n-r columns with no defects, which can be written based on the rule of product as:

$$P(D_1 D_2 \cdots D_r ND_{r+1} \cdots ND_n) = P(D_1 D_2 \cdots D_r) \cdot P(ND_{r+1} \cdots ND_n \mid D_1 D_2 \cdots D_r)$$

The first term probability term is obtained previously, while the second conditional probability is derived from the previous example by substituting N with N-r, M with M-r, and w with n-r, as follows:

$$\begin{aligned} P(D_1 D_2 \cdots D_r ND_{r+1} \cdots ND_n) &= \frac{M!(N-r)!}{(M-r)!N!} \times \frac{(N-M)!(N-r-n+r)!}{(N-M-n+r)!(N-r)!} \\ &= \frac{M!\,(N-M)!\,(N-n)!}{(M-r)!\,N!\,(N-M-n+r)!} \end{aligned}$$

- **What is the probability that exactly *r* columns inspected are with defects? (hint: different configurations of *r* defects in a sample of *n* columns have equal probability, thus, the number of configurations is equal to the Binomial coefficient)**

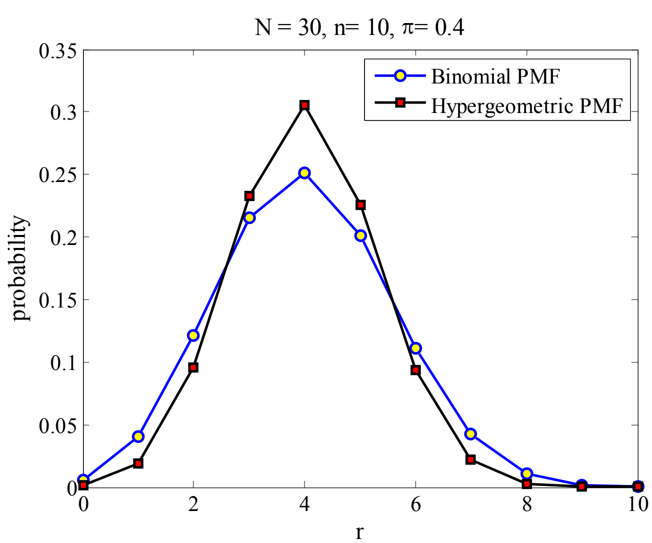
$$P(r \mid N, M, n) = \frac{n!}{r!(n-r)!} \times \frac{M!(N-M)!(N-n)!}{(M-r)!N!(N-M-n+r)!}$$
$$= \frac{M!}{r!(M-r)!} \times \frac{(N-M)!}{(n-r)!(N-M-n+r)!} \times \frac{1}{\frac{N!}{n!(N-n)!}}$$
$$= \frac{\binom{M}{r}\binom{N-M}{n-r}}{\binom{N}{n}}$$

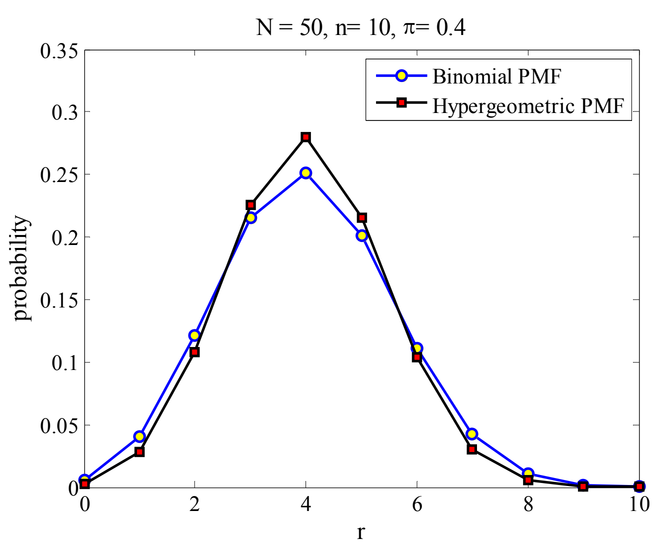
The above distribution is called the *Hypergeometric* distribution.

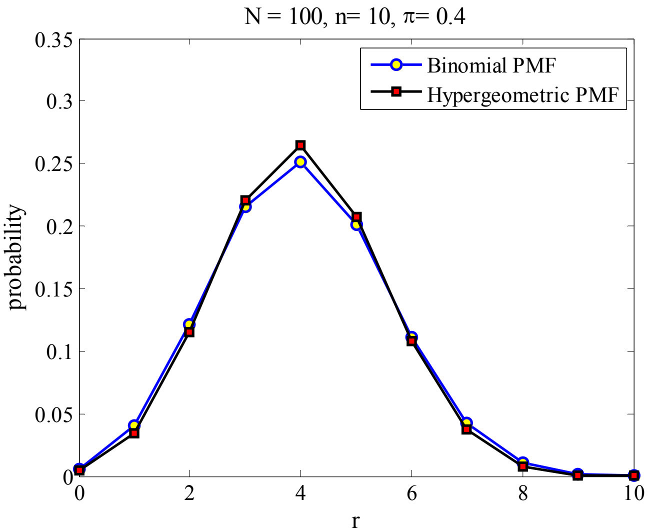
- **Imagine that we have no way of keeping track of which columns we have inspected. In such case, what is the probability of having exactly *r* defects in *n* columns inspected?**

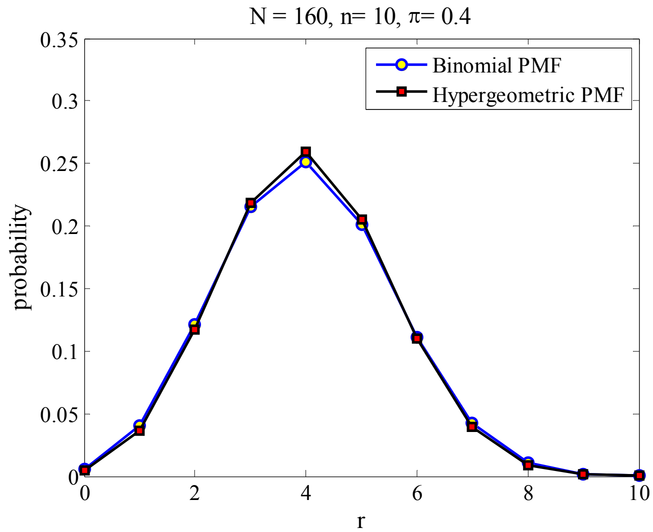
$$P(r \mid N, M, n) = \frac{n!}{r!(n-r)!}\left(\frac{M}{N}\right)^{r}\left(\frac{N-M}{N}\right)^{n-r}$$

Note: if the number N is very large compared to the number n (N>>n), then the hypergeometric distribution approaches in the limit to Binomial distribution. The following examples illustrates this issue assuming that N=[30, 50, 100, 150], n=10, π=0.40, M=0.40N






N = 100, n= 10, π= 0.4
Binomial PMF
Hypergeometric PMF
probability
r


N = 160, n= 10, π= 0.4
Binomial PMF
Hypergeometric PMF
probability
r

5. The European Damage Scales:

The European damage scale classifies the observed structural damage into five distinct levels, as follows:

D1: Negligible to slight damage

D2: Moderate damage

D3: Substantial to heavy damage

D4: Very heavy damage

D5: Destruction

- **In the aftermath of a strong earthquake, if we do not have any specific information what probability you would give to the school remaining without damage (hint: damage level D1)?**

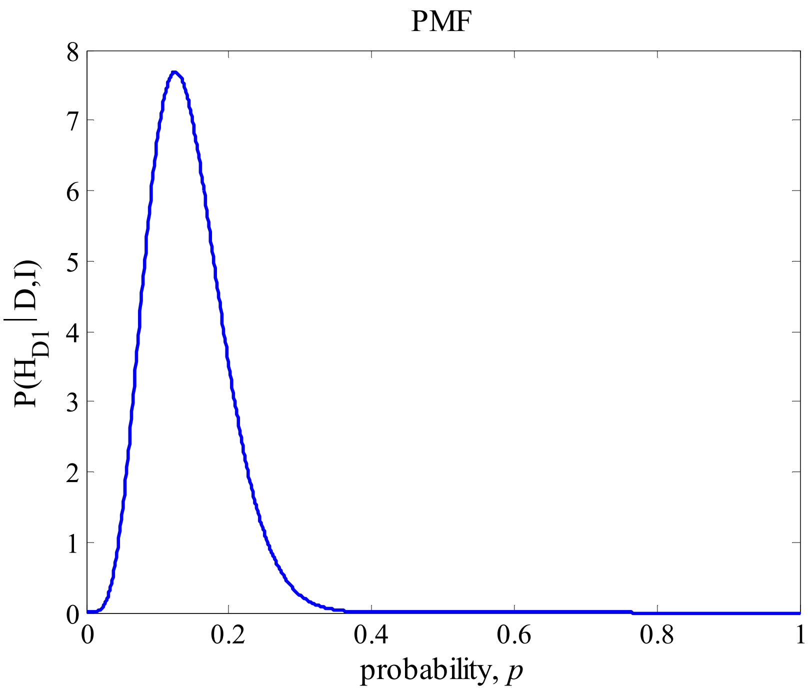
$$P_{\mathrm{D1}} = \frac{1}{5} = 0.20$$

- **Suppose that we have the additional information that out of 40 schools inspected in the aftermath of a strong earthquake, 15 are found to be D2, 10 are found to be between D3, 8 are found to be between D4, 2 are found to be between D5. What is the distribution of the probability that a given school building is intact?**

D: n_{D1}=5 schools are intact out of n=40 inspected schools

H: the probability of having an intact school is equal to p

$$P\left(H_{\mathrm{D1}} \mid D, I\right) == \frac{(n+1)!}{n_{\mathrm{D1}}!\,\left(n-n_{\mathrm{D1}}\right)!}\, p^{n_{\mathrm{D1}}} \left(1-p\right)^{n-n_{\mathrm{D1}}}$$


PMF
8
7
6
5
4
3
2
1
0
0
0.2
0.4
0.6
0.8
1
P(H_D1 | D,I)
probability, p