Chapter 3 – Common Families of Distributions

Section 3.1 - Introduction

Purpose of this Chapter: Catalog many of common statistical distributions (families of distributions that are indexed by one or more parameters)

What we should know about these distributions:

- Definition
- Background
- Descriptive Measures: mean, variance, moment generating function
- Typical applications
- Interesting and useful interrelationships

Section 3.2 – Discrete Distributions

Discrete Uniform Distribution $(1, N)$

Definition: A random variable X has a *discrete uniform* $(1, N)$ distribution if $P(X = x \mid N) = 1/N$, $x = 1, 2, \cdots, N$, where N is a specified positive integer.

Descriptive Measures: $EX = \frac{N+1}{2}$, $VarX = \frac{(N+1)(N-1)}{12}$, $M_X(t) = \frac{1}{N}\sum_{k=1}^{N} e^{kt}$ using the formula $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ and $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

Generalized Discrete Uniform (N_0, N_1): $P(X = x \mid N_0, N_1) = 1/(N_1 - N_0 + 1)$, $x = N_0, \cdots, N_1$. How to calculate EX and $VarX$.

Example: Roll a fair dice, $X =$ the face number, then $X \sim uniform(1,6)$.

Example: Toss a fair coin, define $X = 1$ if we have a head and $X = 0$ otherwise, then $X \sim uniform(0,1)$.

Example: Choose 5 cards from a standard deck of 52 playing cards and consider the unordered sample space S, then each element in S has a probability of $1/\binom{52}{5}$.

Bernoulli Distribution (p)

Definition: A random variable X has a *Bernoulli distribution* with the parameter p if $P(X = 1 \mid p) = p$ and $P(X = 0 \mid p) = 1 - p$, where $0 \le p \le 1$.

Notes:

1. A Bernoulli trial (named after James Bernoulli) is an experiment with two, and only two, possible outcomes.
2. Bernoulli random variable $X = 1$ if "success" occurs and $X = 0$ if "failure" occurs where the probability of a "success" is p.

Descriptive Measures: $EX = p$, $VarX = p(1-p)$, $M_X(t) = pe^t + (1-p)$.

Example: Roll a fair dice, define $X = 1$ if the face number is 1 or 4 and $X = 0$ otherwise, then $X \sim Bernoulli(1/3)$.

Example: Toss a fair coin, define $X = 1$ if we have a head and $X = 0$ otherwise, then $X \sim Bernoulli(1/2)$.

Binomial Distribution (n, p)

Definition: A random variable X has a *binomial distribution* with parameters (n, p) if

$$P(X = x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ where } x = 0, 1, \cdots, n,\ n > 0, \text{ and } 0 \le p \le 1.$$

Notes:

1. A Binomial experiment consists of n independent identical Bernoulli trials, i.e., the experiment consists of a sequence of n trials, where n is fixed in advance of the experiment.
2. The trials are identical, and each trial can result in one of the same two possible outcomes, which we denote by success (S) and failure (F).
3. The trials are independent, so that the outcome on any particular trial does not influence the outcome on any other trial.
4. The probability of success is constant from trial to trial; we denote this probability by p.
5. $X = \sum_{i=1}^{n} Y_i$, where $Y_1, \cdots, Y_n$ are n identical, independent Bernoulli random variables. Hence, the sum of n identical, independent Bernoulli random variables has a Binomial distribution (n, p).

<u>Descriptive Measures:</u> $EX = np$, $VarX = np(1-p)$, $M_X(t) = (pe^t + (1-p))^n$.

<u>Example:</u> A new drug that is effective in 30% percent of its applications is administrated to 10 randomly chosen patients. What is the probability that it will be effective in five or more of these patients?

<u>Solution:</u> Let $Y_i = 1$ if the drug is effective for the ith patient and $Y_i = 0$ otherwise. Let $X = \sum_{i=1}^{10} Y_i$, then $X \sim Binomial(10, 0.3)$. We need to find out that $P(X \geq 5)$ and we have:

$$P(X \geq 5) = \sum_{i=5}^{10} \binom{10}{i} 0.3^i (1-0.3)^{10-i} = 0.1503.$$

Application (Example 3.2.3 - Dice probabilities) Suppose we are interested in finding the probability of obtaining at least on 6 in four rolls of a fair die. This experiment can be modeled as a sequence of four Bernoulli trails with success probability $p=1/6$. Let X =number of 6 in four rolls, then X ~binomial$(4,1/6)$. Thus

$$P(X\ge 1)=1-P(X=0)=1-\binom{4}{0}(\frac{1}{6})^0(\frac{5}{6})^4=0.5177.$$

Example (Estimating p): In some situations, p is unknown and we need to estimate it. To do this, we find p_0 such that the probability value of k is maximum:

$$P(X=k\mid p_0)=\binom{n}{k}p_0^k(1-p_0)^{n-k}=\max_{0\le 1\le p}P(X=k\mid p)=\max_{0\le 1\le p}\binom{n}{k}p^k(1-p)^{n-k}.$$

We notice that:

$$\frac{d}{dp}P(X=k)=\binom{n}{k}[kp^{k-1}(1-p)^{n-k}-(n-k)p^k(1-p)^{n-k-1}]=0.$$

So we have: $k(1-p)-(n-k)p=0$, then $p=k/n$.

In addition, we have $\frac{d^2}{dp^2}P(X=k)<0$, so when $p=k/n$, $P(X=k)$ is the maximum.

As an example, if we observed 10 successes in 30 trials, then our estimate of p would be 10/30=0.333.

Hypergeometric Distribution (N,M,K)

Example (The capture-recapture procedure to estimate the size of an animal population): A sample size of M is captured, tagged, and then released into the population. After awhile a new catch of K animals is made, and the number of tagged animals in the sample is noted. Let X be the number of tagged animals in the second catch. It provides an estimate of total population of animals, N. For given N, M, K, what is the distribution of X? If we observed $M = 100$, $K = 50$, and $X = 10$, what is the estimate of N?

Definition: A discrete random variable X has the *hypergeometric distribution* with parameters (N, M, K) if

$$P(X = x \mid N, M, K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}, \max(0, M-(N-K)) \le x \le \min(M, K).$$

Assumptions of the hypergeometric distribution:

1. The population or set to be sampled consists of N individuals, objects, or elements (or a finite population).
2. Each individual can be characterized as a success (S) or failure (F), and there are M successes in the population.
3. A sample of K individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen, i.e., obtain a random sample.
4. Number of success is the random variable.

Hypergeometric random variable: X = the number of success among the K trials

Descriptive Measures: $EX = K\frac{M}{N}$, $VarX = K\frac{M}{N}(\frac{(N-M)(N-K)}{N(N-1)}) = K\frac{M}{N}(1-\frac{M}{N})(\frac{N-K}{N-1})$.

Example (Differences between Binomial and Hypergeometric): We consider a box containing M defectives and $N-M$ nondefectives. We take a sample size of K from the box. And let X be the number of defectives, then the distribution of X depends on whether we sample with or without replacement.

Solution:

If we sample with replacement and consider each draw a trial and identify success as a defective item drawn, then the trials are independent, dichotomous, and the probability of success on each trial is $p = M/N$. Therefore,

$$P(X=x) = \binom{K}{x}(\frac{M}{N})^x(1-\frac{M}{N})^{K-x}, x = 0,1,\cdots,K.$$

If we sample without replacement, then:

$$P(X=x) = \binom{M}{x}\binom{N-M}{K-x}/\binom{N}{K}, \max(0, M-(N-K)) \le x \le \min(M,K).$$

If we sample without replacement, then we can either sample one after another or make a single draw of size K. Two procedures are identical but dependent and we can still show that the probability of success on any draw is $p = M/N$. To prove this, define the event $A_x = \{\text{Success on the } x\text{th draw}\}$ ($x = 1,2,\cdots,K$). Then A_x happens if and only if the xth draw is a success and in the first $x-1$ draws there are j defectives and $x-1-j$ nondefectives, $j = 0,1,\cdots,\min(x-1, M-1)$. Hence

$$P(A_x)=\sum_{j=0}^{\min(x-1,M-1)}\frac{\binom{M}{j}\binom{N-M}{x-1-j}}{\binom{N}{x-1}}\frac{M-j}{N-x+1}=\frac{M}{N}\sum_{j=0}^{\min(x-1,M-1)}\frac{\binom{M-1}{j}\binom{N-M}{x-1-j}}{\binom{N-1}{x-1}}=\frac{M}{N}.$$

If we use $K=2$ as an example, then $P(A_1)=M/N$, $P(A_2\mid A_1)=\frac{M-1}{N-1}$, and

$$P(A_2)=P(A_2\cap A_1)+P(A_2\bigcap A_1^c)=P(A_2\mid A_1)P(A_1)+P(A_2\mid A_1^c)P(A_1^c)=\frac{M-1}{N-1}\frac{M}{N}+\frac{M}{N-1}(1-\frac{M}{N})=\frac{M}{N}.$$

<u>In summary,</u> we have:

	Binomial	**Hypergeometric**
1	Dichotomous trials	Dichotomous trials
2	Trials independent (sampling with replacement)	Trials dependent (sampling without replacement)
3	Probability of success same on any trial	Probability of success same on any trial
4	Number of trials is K	Number of trials is K
5	Infinite population	Finite population

<u>Note:</u> $\frac{N-K}{N-1}$ is known as the finite population correction factor which is always less than one. Hence, the hypergeometric distribution has smaller variance than does the binomial random variable.

Note: There is an easy way to calculate the mean of the hypergeometric distribution. Let $Y_i = 1$ if the i th trial is success and $Y_i = 0$ if the i th trial is failure. Then $X = \sum_{i=1}^{K} Y_i$ and $P(Y_i = 1) = \frac{M}{N}$. In the next chapter (multiple random variables), we will prove that $EX = \sum_{i=1}^{K} EY_i$. Then $EX = K\frac{M}{N}$. In addition, the calculation of $VarX$ can also be simplified by the similar way.

Note: $\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \to \binom{K}{x} p^x (1-p)^{K-x}$ for $M/N \to p$ when $N \to \infty$. From experience, we can use this approximation when $N \geq 50$ and $M/N \leq 0.10$.

Application (Example 3.2.1 - Accepting sampling) Define N =number of item, M =number of defective items. Then we can calculate the probability that a sample size K contains x defective items. For example, if $N = 100$, $M = 6$, $K = 10$, $x = 0$, then $P(X = 0) = \frac{\binom{6}{0}\binom{100-6}{10-0}}{\binom{100}{10}} = 0.5223$.

If we use the Binomial distribution as an approximation, we have

$$P(X=0)=0.06^0(1-0.06)^{10}=0.5386.$$

If $N=1000,$ we have $P(X=0)=0.9416$ (Hypergeometric) and $P(X=0)=0.9417$ (Binomial).

Application (Continue) Suppose $N=100$. The lot will be accepted if $M\le 5$. We randomly check K parts and accept them if and only if we find there is no defective item. What is the minimum K that can ensure that we accept an unacceptable lot with a probability less than 0.10?

Solution: $P(X=0\,|\,M,K)=\dfrac{\binom{M}{0}\binom{100-M}{K}}{\binom{100}{K}}\le 0.10$ for $M>5$. However, we do not know M. We need to calculate this probability for $0\le M\le 100$. To find the minimum K, we only need to calculate $M=6$. We can get $K=32$. $P(X=0\,|\,K=31)=0.1005$ and $P(X=0\,|\,K=32)=0.0918$.

If we use the Binomial distribution as an approximation, we have

$$P(X=0\,|\,p=6/100,K=37)=0.1013 \text{ and } P(X=0\,|\,p=6/100,K=38)=0.0952,$$

so we get $K=38$.

If $N=1000$, we can get $K=320$ ($P(X=0\,|\,K=319)=0.1007$ and $P(X=0\,|\,K=320)=0.099$ for Hypergeometric) and $K=383$ ($P(X=0\,|\,K=382)=0.1007$ and $P(X=0\,|\,K=383)=0.0998$ for Binomial).

Poisson Distribution (λ)

Definition: A random variable X has a *Poisson distribution* with the parameter $(\lambda>0)$ if

$$P(X = x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,\cdots.$$

Notes:

1. A Poisson distribution is typically used to model the probability distribution of the number of occurrences (with λ being the intensity rate) per unit time or per unit area.
2. A basic assumption: for small time intervals, the probability of an event occurring is proportional to the length of waiting time.
3. It was shown in Section 2.3 that Binomial pmf approximates Poisson pmf. Poisson pmf is also a limiting distribution of a negative binomial distribution.
4. A useful result: By Taylor series expansion, we have $e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$.

Note: Let $X(\Delta)$ be the number of events that occur during an interval Δ.The Poisson distribution can be derived based on following assumptions:

1. The events are independent; this means if $\Delta_1, \cdots, \Delta_n$ are disjoint intervals then $X(\Delta_1), \cdots, X(\Delta_n)$ are independent.
2. The distribution of $X(\Delta)$ depends only on the length of Δ and not on the time of occurrence.
3. The probability that exactly one event occurs in a small interval of length Δt equals $\lambda\Delta t + o(\Delta t)$ where $o(\Delta t)/\Delta t \to 0$ when $\Delta t \to 0$.
4. The probability that two or more events occur in a small interval of length Δt is $o(\Delta t)$.

Descriptive Measures: $EX = \lambda$, $VarX = \lambda$, $M_X(t) = e^{\lambda(e^t - 1)}$

Application (Example 3.2.4 – Waiting Time) If there are five calls in 3 minutes in average, what is the probability that there will be no calls in the next minute?

Solution: Let $X =$number of calls in a minute, then X has a Poisson distribution with $EX = \lambda = 5/3$, thus

$P(X = 0) = \dfrac{e^{-5/3}(5/3)^0}{0!} = 0.189$, $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 0.496.$

Application (Example 3.2.5 – Poisson approximation) A typesetter, on the average, makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be not more than two errors in five pages.

Solution: Let $X =$number of errors in five pages, then $X \sim \text{binomial}(n, p)$, where $n = 1500$ and $p = 1/500$. Thus

$$P(X \leq 2) = \sum_{x=0}^{2} \binom{1500}{x} (\frac{1}{500})^x (1 - \frac{1}{500})^{1500-x} = 0.4230.$$

If we use the Poisson approximation, we have $\lambda = np = 3$, and $P(X \leq 2) \approx e^{-3}(1 + 3 + 3^2/2) = 0.4232$.

Negative Binomial Distribution (r, p)

Definition: A random variable X has a *negative binomial distribution* with parameters (r, p) if

$$P(X = x \mid r, p) = \binom{r + x - 1}{x} p^r (1 - p)^x, x = 0, 1, \cdots.$$

Assumptions and Notes:

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in either a success (S) or failure (F).
3. The probability of success, p, is constant from trial to trial.
4. The experiment continues (trials are performed) until a total of r successes have been observed, where r is a specified positive integer.
5. In contrast to the binomial experiment where the number of trials is fixed and the number of successes is random, the negative binomial experiment has the number of successes fixed and the number of trials random.
6. Define $X =$ total number of failures until the r th success. Then $Y = X + r$ is the total number of trials until the r th success and X has a negative binomial distribution. Then,
$$P(Y = y) = P(X = y - r) = \binom{y-1}{r-1} p^r (1-p)^{y-r} \, (y \ge r).$$
7. We have $P(X = x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x = (-1)^x \binom{-r}{x} p^r (1-p)^y, x = 0,1,\cdots$. hence the name "negative binomial".
8. It is not easy to verify that $\sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r (1-p)^{x-r} = 1$.

Descriptive Measures: $EX = \frac{r(1-p)}{p}$, $VarX = \frac{r(1-p)}{p^2}$, $M_X(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r$, then $EY = \frac{r(1-p)}{p} + r$, $VarY = \frac{r(1-p)}{p^2}$, $M_Y(t) = e^{rt}\left[\frac{p}{1-(1-p)e^t}\right]^r$. How about $EY, VarY, M_Y(t)$?

Solution: To calculate all the measures, we first recall the Taylor series expansion for $g(x) = (1-x)^{-r}$. According to the Taylor series expansion $g(x) = \sum_{k=0}^{\infty} \frac{d}{dx^k} g(x)|_{x=0} \frac{x^k}{k!}$.

We have $\frac{d}{dx^k}(1-x)^{-r} = (-1)^k(-r)(-r-1)\cdots(-r-k+1)(1-x)^{-r-k}$.

Therefore , $g(x) = \sum_{k=0}^{\infty} (-1)^k \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!} x^k = \sum_{k=0}^{\infty} \binom{r+k-1}{k} x^k$. Then we have

(1) $\sum_{x=0}^{\infty} \binom{r+x-1}{x} p^r (1-p)^x = p^r(1-(1-p))^{-r} = 1.$

(2) $EX = \sum_{x=0}^{\infty} x \binom{r+x-1}{x} p^r (1-p)^x = rp^{-1}(1-p)\sum_{x=1}^{\infty} \binom{r+x-1}{x-1} p^{r+1}(1-p)^{x-1} = \frac{r(1-p)}{p}.$

(3) $M_X(t) = Ee^{Xt} = \sum_{x=0}^{\infty} \binom{r+x-1}{x} p^r e^{xt}(1-p)^x = p^r \sum_{x=0}^{\infty} \binom{r+x-1}{x}((1-p)e^t)^x = \frac{p^r}{(1-(1-p)e^t)^r}$

(4) If $\lim_{r\to\infty} r(1-p) = \lambda$ and $p \to 1$, then $EX = \frac{r(1-p)}{p} \to \lambda$, $VarX = \frac{r(1-p)}{p^2} \to \lambda$. In addition,

$$M_X(t) = [\frac{p}{1-(1-p)e^t}]^r = (1+\frac{p-1+(1-p)e^t}{1-(1-p)e^t})^r = (1+\frac{(e^t-1)(1-p)}{1-(1-p)e^t})^r \to e^{\lambda(e^t-1)}.$$

Application (Example 3.2.6 – Inverse binomial sampling) A technique known as inverse binomial sampling is useful in sampling biological populations. If the proportion of individuals with certain characteristics is p and we sample until we see r such individuals, then the number of individuals sampled is a negative binomial random variable.

Example (Selecting Mice for a Laboratory Experiment): In many laboratory experiments involving animals, the experimenter needs a certain number of animals. Suppose an experimenter needs five mice that have a certain disease. A large collection of mice is available to him, and the incidence of disease in this population is 30%. If the experimenter examines mice one at a time until he gets the five mice needed, then the number of mice X that must be examined has a negative binomial distribution with $p = 0.3$ and $r = 5$. The expected number of examinations required is $r/p = 16.67$. The probability that the number if examinations required is more than 25 is given by $P(X > 25) = 0.0905$, so there is only about 9% chance that more than 25 examinations will be required to get five mice that have the disease.

Question: Suppose that we have N mice and M with the disease, then what is the exact distribution of X, then number of examinations, if we want to get r mice with the disease?

Solution: $P(X = x) = \frac{\binom{M}{r-1}\binom{N-M}{(x-1)-(r-1)}}{\binom{N}{x-1}} \frac{M-(r-1)}{N-(x-1)}, x = r, r+1, \cdots, N-M+r.$ Rewrite this formula, we have:

$$P(X = x) = \binom{x-1}{r-1} \frac{\binom{N-x}{M-r}}{\binom{N}{M}}.$$

Comparing with the sample with replacement, we have:

$$P(X = x) = \binom{x-1}{r-1} \left(\frac{M}{N}\right)^r \left(1 - \frac{M}{N}\right)^{x-r}.$$

Geometric Distribution (p)

Definition: A random variable X has a *geometric distribution* with parameters (p) if

$$P(X = x \mid p) = p(1-p)^{x-1}, x = 1, 2, \cdots.$$

Assumptions and Notes:

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in either a success (S) or failure (F).
3. The probability of success, p, is constant from trial to trial.
4. The experiment continues (trials are performed) until the first success.
5. The geometric distribution is a special case of the negative binomial with r = 1.

Descriptive Measures: $EX = \frac{1-p}{p} + 1 = \frac{1}{p}$, $VarX = \frac{1-p}{p^2}$, $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$.

Memoryless property of the Geometric Distribution:

$$P(X > s \mid X > t) = P(X > s - t) \text{ for } s > t.$$

Proof: $P(X > s \mid X > t) = P(X > s)/P(X > t) = (1-p)^s/(1-p)^t = (1-p)^{s-t}$.

Application (Example 3.2.7 – Failure times) The geometric distribution is sometimes used to model “lifetimes” or “time until failure” of components.

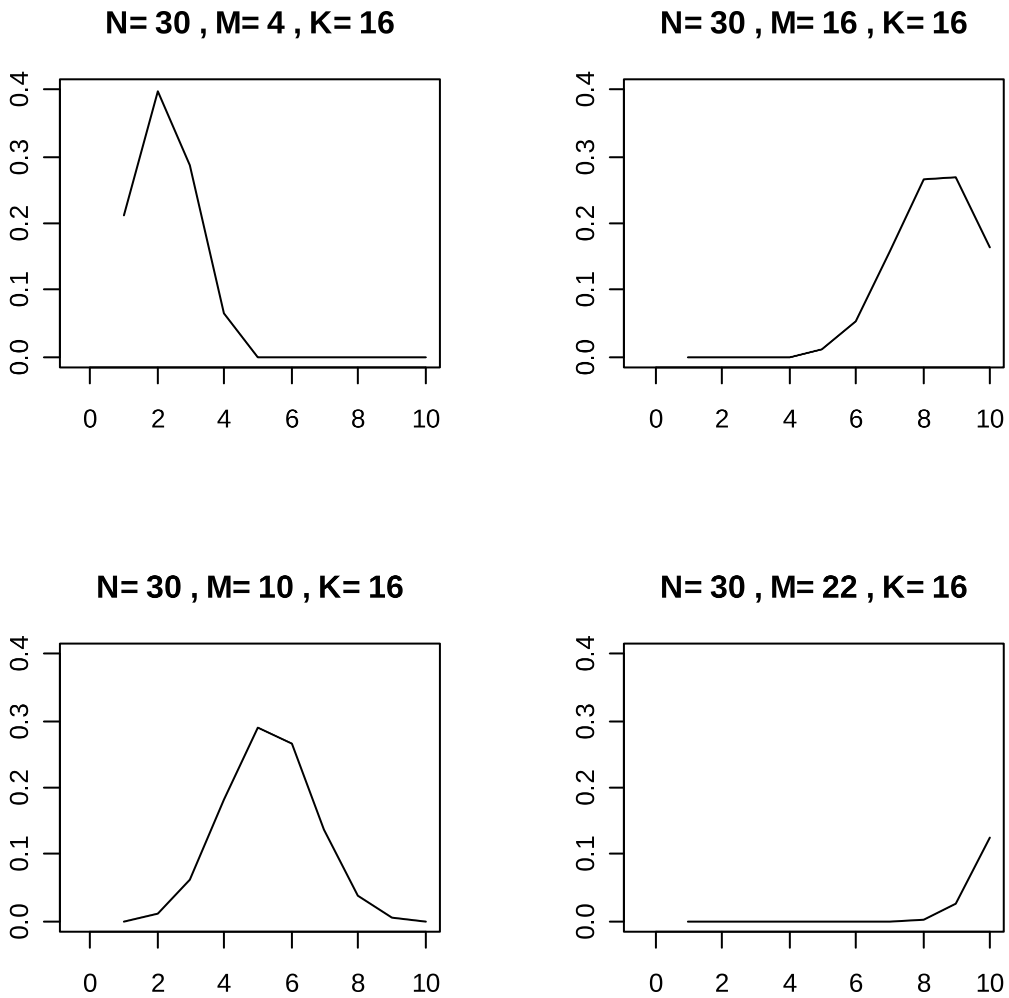

Figure 1. The pmf of Hypergeometric distribution with different parameters (N, M, K).

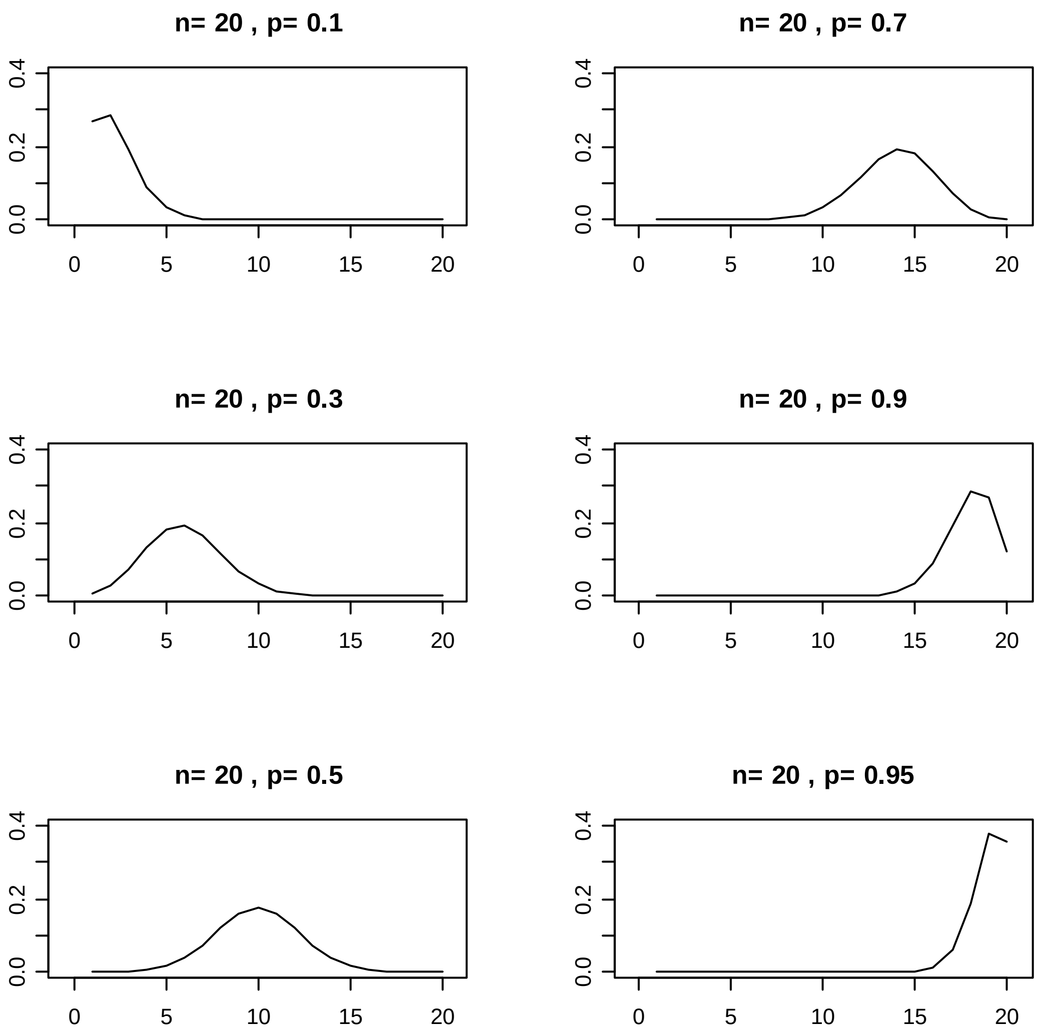

Figure 2. The pmf of Binomial distribution with different parameters (n, p).

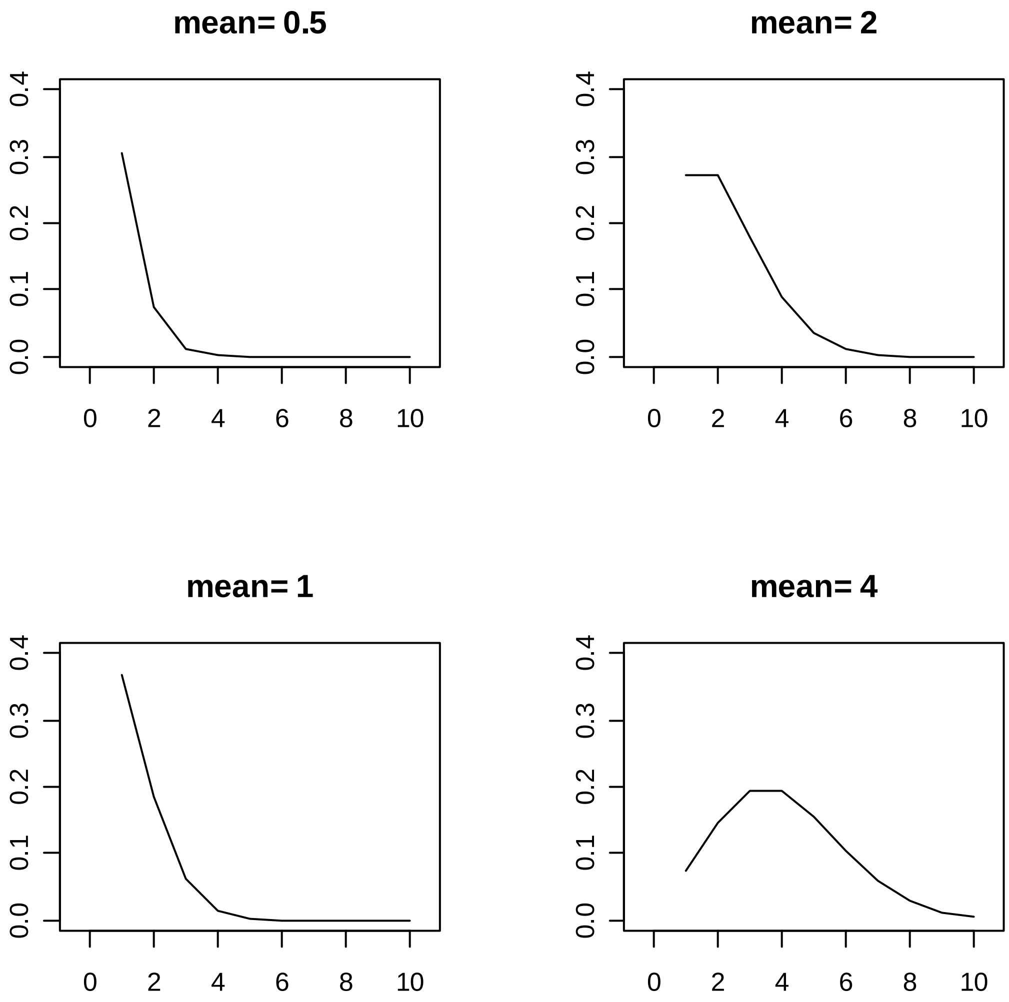

Figure 3. The pmf of Poisson distribution with different parameters λ.

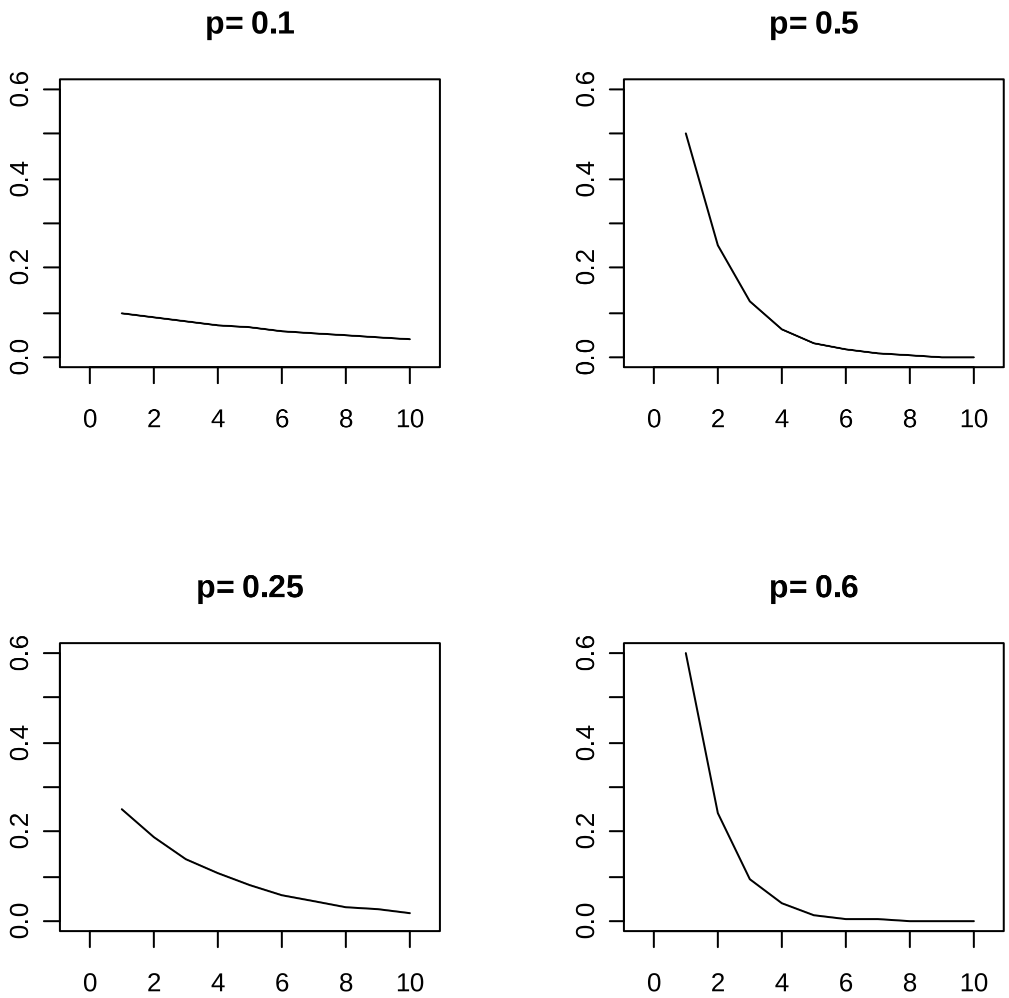

Figure 4. The pmf of Geometric distribution with different parameters p.