Handout 2

Binomial and Multinomial Distributions.

Binomial distributions come up when we deal with binary (categorical) variables. In Example 2 of Handout 1, we discussed counting the number of graduates in a random sample of $n = 200$ adults in a large city. Consider binary random variables $Y_1, \ldots, Y_n$, where each Y_i takes two values: 1 (college graduate), 0 (high school). Then each Y_i is called a Bernoulli random variable and $Y = Y_1 + \cdots + Y_n$ is called a binomial random variable. Note that the range of possible values of Y is $0, \ldots, n$. Let π be the proportion of college graduates in the city. Thus we can say that we have a coin with the faces marked S (success, college graduate) and F (failure, not a college graduate), and the probability of S is π. We toss this coin n times. If the i^{th} toss results in S, then Y_i is 1, and is 0 otherwise. Then $Y = Y_1 + \cdots + Y_n$ is the total number of S's out of n tosses. In order for Y to be called a binomial random variable, the following assumptions are essential:

(i) each Y_i can assume only two values: 0 or 1,

(i) $P(Y_i = 1) = \pi$ and $P(Y_i = 0) = 1 - \pi$,

(ii) probability of S on any toss is the same,

(iii) $Y_1, \ldots, Y_n$ are independent random variables.

The probability distribution is usually denoted by $Binomial(n, \pi)$, and we write $Y \sim Binomial(n, \pi)$.

Note that

$$\begin{aligned}
E(Y_i) &= (0)P(Y_i = 0) + (1)P(Y_i = 1) = (0)(1-\pi) + (1)(\pi) = \pi, \\
Var(Y_i^2) &= (0^2)P(Y_i = 0) + (1^2)P(Y_i = 1) = (0)(1-\pi) + (1)(\pi) = \pi, \\
Var(Y_i) &= E(Y_i^2) - [E(Y_i)]^2 = \pi - \pi^2 = \pi(1-\pi), \\
E(Y) &= E(Y_1) + \cdots + E(Y_n) = n\pi, \\
Var(Y) &= Var(Y_1) + \cdots + Var(Y_n) \quad [Y_1, \ldots, Y_n \text{ are independent}] \\
&= n\pi(1-\pi).
\end{aligned}$$

Probability density (or mass) function of Y.

Let $Y \sim Binomial(n, \pi)$, then its probability density function is given by

$$P(Y = y) = \binom{n}{y} \pi^y (1-\pi)^{n-y} = \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y}, \; y = 0, \ldots, n.$$

Clearly, $\sum_{y=0}^{n} P(Y = y) = 1$.

Example 1. Suppose that a 20% of the adults in a state are college graduates. A random sample on $n = 10$ adults is taken. Let Y be the number of college graduate in the sample. Then $Y \sim Binomial(10, 0.2)$. Chance that the sample contains at least 2 college graduates is given by

$$\begin{aligned}
P(Y &\geq 2) = 1 - P(Y \leq 1) \\
&= 1 - [P(Y = 0) + P(Y = 1)] \\
&= 1 - \left[\binom{10}{0} (0.2)^0 (0.8)^{10} + \binom{10}{1} (0.2)^1 (0.8)^9 \right] \\
&= 1 - [0.1074 + 0.2684] = 0.6242.
\end{aligned}$$

Note that

$$\begin{aligned} E(Y) &= n\pi = (10)(.2) = 2, \\ Var(Y) &= n\pi(1-\pi) = (10)(0.2)(0.8) = 1.6. \end{aligned}$$

Central Limit Theorem.

If $X_1, \ldots, X_n$ are independent and identically distributed (iid) with mean μ and variance σ^2. Then the sample mean $\bar{X} = (X_1 + \cdots X_n)/n$ and the sample total $S = n\bar{X} = X_1 + \cdots + X_n$ have the following means and variances

$$\begin{aligned} E(\bar{X}) &= \mu, Var(\bar{X}) = \sigma^2/n, \\ E(S) &= n\mu, Var(S) = n\sigma^2. \end{aligned}$$

A fundamental result from probability is that when n is large, the probability distribution of $\bar{X}$ is approximately $N(\mu, \sigma^2/n)$, or equivalently, the probability distribution of S is approximately $N(n\mu, n\sigma^2)$. How large does n need to be in order for this approximation to be reasonably good? There is a clear answer to that. If X_i normally distributed, then the distribution of $\bar{X}$ (or S) is exactly normal for any n. Farther away the distribution of X_i from normality, the larger the value of n is needed for the normal approximation for the probability distribution of $\bar{X}$.

If $Y \sim Binomial(n, \pi)$, then the central limit theorem holds for Y since Y is a sum of iid Bernoulli variables with $E(Y_i) = \pi$ and $Var(Y_i) = \pi(1-\pi)$. Thus for n large, the distribution of Y is approximately $N(n\pi, n\pi(1-\pi))$. The graphs given later shows the probability distributions of Y for $n = 10$, $n = 25$ and $n = 50$ when $\pi = 0.1$ and when $\pi = 0.5$.

When $\pi = 0.5$, the distribution of Y is not too far away for normality even when $n = 10$. But when $\pi = 0.1$, the distribution of Y is not close to normality even when $n = 25$. The reason is that the distribution of Y_i is symmetric when $\pi = 0.5$, but the distribution of Y_i is very skewed when $\pi = 0.1$.

Mathematically, normal approximation to the binomial distribution requires that both $n\pi$ and $n(1-\pi)$ are large. As a rule of thumb, it is often said that we may use normal approximation for the binomial if both $n\pi$ and $n(1-\pi)$ are 5 or larger.

Estimation for Binomial

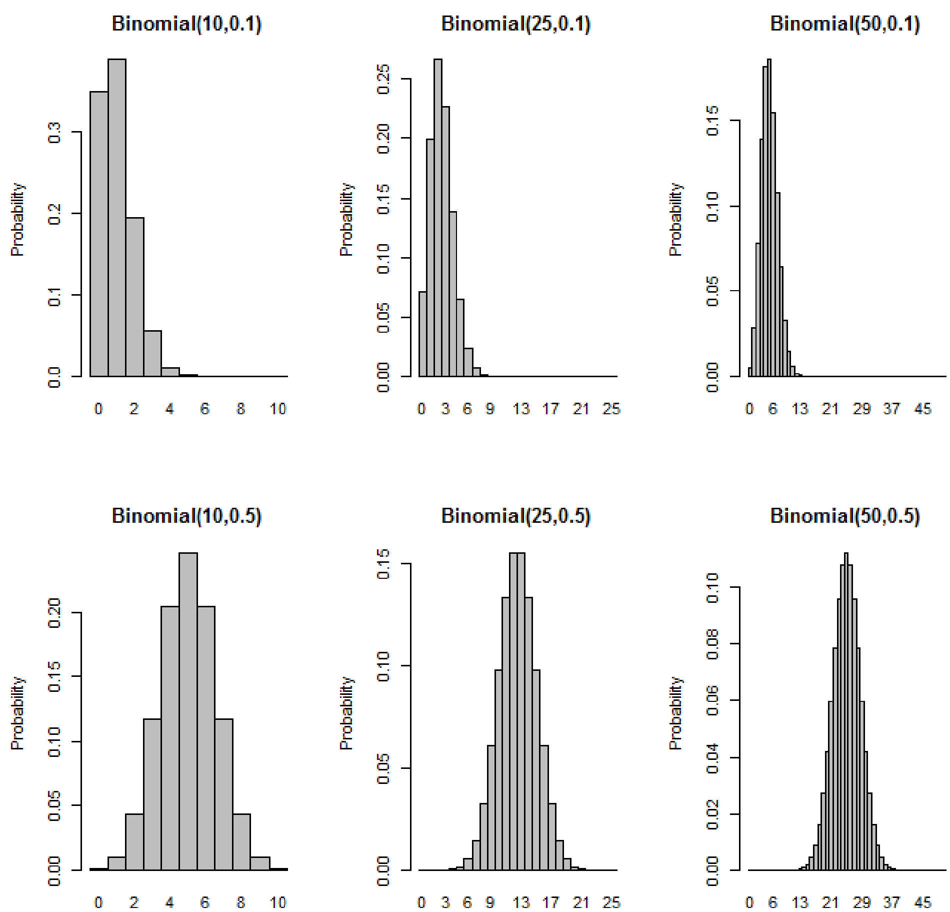
We would like to estimate the proportion π of college graduate in a city. Let Y be the number of college graduate in a random sample of size n. The maximum likelihood (to be described later) estimator of π is given $p = Y/n$. The formula for the mean and variance of Y tells us

$$E(p) = \pi \text{ and } Var(p) = \pi(1-\pi)/n.$$

Assuming both $n\pi$ and $n(1-\pi)$ to be large, p is approximately $N(\pi, \pi(1-\pi)/n)$, and thus $Z = (p - \pi)/\sqrt{\pi(1-\pi)/n}$ is approximate $N(0,1)$. So an approximately 95% confidence interval of π is given by $p \pm 1.96SE(p)$, where $SE(p) = \sqrt{p(1-p)/n}$.

Assume that in a random sample of $n = 200$, we have found that 71 are college graduates. Thus $p = 71/200 = 0.355$ is an estimate of π. Note that

$$SE(p) = \sqrt{p(1-p)/n} = \sqrt{(0.355)(1-0.355)/200} = 0.03384$$


Binomial(10,0.1)
Binomial(25,0.1)
Binomial(50,0.1)
Binomial(10,0.5)
Binomial(25,0.5)
Binomial(50,0.5)
Probability
0 2 4 6 8 10
0 3 6 9 13 17 21 25
0 6 13 21 29 37 45

So an approximate 95% confidence interval for π is given by

$$
\begin{aligned}
& p \pm 1.96 SE(p), ie, 0.355 \pm (1.96)(0.03384), \\
& ie, 0.355 \pm 0.0663, ie, (0.2887, 0.4213).
\end{aligned}
$$

Multinomial distribution

As part of a survey on health care, a random sample of $n = 125$ adults is taken in large city, and each person in the sample is asked whether he/she is satisfied or unsatisfied or undecided with the current healthcare system. Note that the categorical variable is 'opinion on the health care system' which has 3 categories. For the i^{th} person, we thus have

$$Y_{i1} = \begin{cases} 1 & \text{satisfied} \\ 0 & \text{otherwise} \end{cases}, Y_{i2} == \begin{cases} 1 & \text{unsatisfied} \\ 0 & \text{otherwise} \end{cases}, Y_{i3} = \begin{cases} 1 & \text{undecided} \\ 0 & \text{otherwise} \end{cases}.$$

Thus the vector variable $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})$ constitutes the opinion of the i^{th} person and $Y_{i1}+Y_{i2}+Y_{i3} = 1$. If there are c categories, the vector Y_i has c components $Y_{i1}, \ldots, Y_{ic}$, and $Y_{i1} + \cdots + Y_{ic} = 1$.

Let

$$
\begin{aligned}
n_1 &= Y_{11} + \cdots Y_{n1} = \sum_{i=1}^{n} Y_{i1}, \\
&\vdots \\
n_c &= Y_{1c} + \cdots + Y_{nc} = \sum_{i=1}^{n} Y_{ic}.
\end{aligned}
$$

So n_1 is the number of adults in the sample who are satisfied, n_2 is the number of 'unsatisfied' in the sample and so on. Note that $n_1 + \cdots + n_c = n = 125$.

Let π_1 be the proportion of adults in the city who are satisfied, π_2 be the proportion who are unsatisfied and so on. In that case $\pi_1 + \cdots + \pi_c = 1$, and for each $i = 1, \ldots, n$,

$$P(Y_{ij} = 1) = \pi_j, j = 1, \ldots, c.$$

The vector of counts $(n_1, \ldots, n_c)$ is said to have a multinomial distribution (and denoted by $(n_1, \ldots, n_c) \sim Multinomial(n, \pi_1, \ldots, \pi_c)$) whose probability density (or mass) function is given by

$$p(n_1, \ldots, n_c) = \frac{n!}{n_1! \cdots n_c!} \pi_1^{n_1} \cdots \pi_c^{n_c}, \text{ with } n_1 + \cdots + n_c = n..$$

Here are some basic results on the multinomial distribution.

Fact 1. Some properties of the multinomial distribution.

(a) For each j, $n_j \sim Bionomial(n, \pi_j)$,

(b) $E(n_j) = n\pi_j, Var(n_j) = n\pi_j(1 - \pi_j)$,

(c) $Cov(n_j, n_k) = -n\pi_j\pi_k$, $j \neq k$.

Example 2. Assume that in a large city, 50% are satisfied with the current health care system, 30% are unsatisfied, and the rest are undecided. A random sample of $n = 125$ residents will be taken. Let n_1, n_2, n_3 be the number of individuals in the sample who are 'satisfied', 'unsatisfied' and 'undecided' respectively. Note that here $c = 3$, $\pi_1 = 0.50, \pi_2 = 0.3, \pi_3 = 0.2$. The probability that there are 63 satisfied, 38 unsatisfied and 24 undecided in the sample is given by

$$\begin{aligned} & \frac{n!}{n_1! n_2! n_3!} \pi_1^{n_1} \pi_2^{n_2} \pi_3^{n_3} \\ = & \frac{125!}{63!38!24!} (0.5)^{63} (0.3)^{38} (0.2)^{24} = 0.0072. \end{aligned}$$

We also have

$$\begin{aligned} E(n_1) &= n\pi_1 = (125)(0.50) = 62.5, \\ E(n_2) &= n\pi_2 = (125)(0.3) = 37.5, \\ E(n_3) &= n\pi_3 = (125)(0.2) = 25, \\ Var(n_1) &= n\pi_1(1-\pi_1) = (125)(0.5)(1-0.5) = 31.25, \\ Var(n_2) &= n\pi_2(1-\pi_2) = (125)(0.3)(1-0.3) = 26.25, \\ Var(n_3) &= n\pi_3(1-\pi_3) = (125)(0.2)(1-0.2) = 20, \\ Cov(n_1, n_2) &= -n\pi_1\pi_2 = -(125)(0.5)(0.3) = -18.75, \\ Cov(n_1, n_3) &= -n\pi_1\pi_3 = -(125)(0.5)(0.2) = -12.5, \\ Cov(n_2, n_3) &= -n\pi_2\pi_3 = -(125)(0.3)(0.2) = -7.5, \end{aligned}$$

Estimation for Multinomial

If $(n_1, \ldots, n_c) \sim Multinomial(n, \pi_1, \ldots, \pi_c)$, then the MLE of π_j is given by $p_j = n_j/n$, $j = 1, \ldots, c$. From the mean, variance and covariance formulas for n_j's, we can deduce that

$$\begin{aligned} E(p_j) &= \pi_j,\ Var(p_j) = \pi_j(1-\pi_j)/n, \text{ and} \\ Cov(p_j, p_k) &= -\pi_j\pi_k/n. \end{aligned}$$

Note that p_j's are mutually correlated. Since each n_j is binomially distributed and p_j depends only on n_j and n, the CLT (Central Limit Theorem) holds for p_j. Moreover, the distribution of $(p_1, \ldots, p_c)$ is approximately a multivariate normal with the mean vector $(\pi_1, \ldots, \pi_c)$' and a covariance matrix Σ assuming that $n\pi_j$ is large for all j. The j^{th} diagonal element of Σ is $\pi_j(1-\pi_j)/n$, and element (j,k), $j \neq k$, of Σ is $-\pi_j\pi_k/n$. One can construct confidence intervals for π_j as in binomial case.