Handout 3

Poisson and Negative Binomial Distributions.

Poisson and Negative Bionimal distributions often come up in the analysis of categorical data. We will discuss them briefly in this handout.

Poisson Distribution.

There are many real life counts that can be modeled by a Poisson distribution: # of phone calls a certain office receives in the morning in a day, # accidents on Highway 80 between Davis and Sacramento in a week etc. A discrete random variable Y is said to have Poisson distribution with mean μ (and we write $Y \sim Poisson(\mu)$) if its pdf is given by

$$f(y;\mu) = P(Y = y) = \exp(-\mu)\frac{\mu^y}{y!},\ y = 0, 1, 2 \ldots.$$

Note that for a $binomial(n, p)$, the largest value the random variable can take is n. However, in tthe case of a Poisson random variable, there is no such upper limit. Here are some properties of a Poisson random variable.

Fact 1. Let $Y \sim Poisson(\mu)$. Then, $E(Y) = \mu$ and $Var(Y) = \mu$.

Note that the variance equals the mean for a Poisson distribution. The followoing example presents a real data set where a Poisson distribution seems to be appropriate.

Example 1. (Flying-bomb hits on London) During World War II, London was hit by flying bombs. Can we model the # of hits in a particular geographical area by a Poissin distribution? In order to analyze the data, South of London was divided in $N = 576$ small areas of 1/4 square kilometers each, and the table below records the number of areas with 0 hit, 1 hit, 2 hits etc. Since there were a total of 537 hits, average number of hits per 1/4 square kilometer area is $\hat{\mu} = 537/576 = 0.9323$. The table gives observed counts and proportions as well as the (expected) counts and probabilities under $Poisson(0.9323)$.

The fit of the Poisson distribution is surprisingly good; as judged by the χ^2-criterion (to be discussed later). This data was analyzed by Bristish statistician R. D. Clarke.

# of hits	# of areas	Sample Proportion	Poisson Prob	Poisson Count
0	229	0.398	0.394	226.7
1	211	0.366	0.367	211.4
2	93	0.162	0.171	98.5
3	35	0.071	0.053	30.6
4	7	0.012	0.012	7.1
≥ 5	1	0.002	0.003	1.6

It is interesting to note that most people believed in a tendency of the points of impact to cluster. If this were true, there would be a higher frequency of areas with either many hits or no hit and a deficiency in the intermediate classes. The table indicates perfect randomness and homogeneity of the area; we have here an instructive illustration of the established fact that to the untrained eye randomness appears as regularity or tendency to cluster.

Negative Binomial Distribution.

For the Poisson distribution, the mean equal variance. However, there are many cases where the variance is higher than the mean (this phenomenon is called 'overdispersion'). To understand this, consider the variable: # of accidents on Highway 80 between San Francisco and Sacramento in a day (say Wednesday). Now we can think of this section of the highway as composed of 15 subsegments; near Sacramento, near Davis, near Dixon etc. Daily number of accidents in each subsection may be a Poisson variable, but the accident proneness (ie, the means) may be different. Thus the total number of accidents on Highway 80 between Sacramento and San Francisco is the sum of 15 Poisson variables with perhaps different means. This leads to what is known as a mixture of Poisson distributions. There are many different type of mixtures, and one such mixture is known as the negative binomial.

A discrete random variable Y is said to have a negative binomial distribution with parameters μ and k if its probability density function is given by

$$f(y; k, \mu) = P(Y = y) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k}{\mu+k}\right)^k \left(1 - \frac{k}{\mu+k}\right)^y, \ y = 0, 1, 2, \ldots,$$

where $k > 0$ and $\mu > 0$, and $\Gamma(\cdot)$ is the gamma function. This distribution results when, given the mean μ (ie, conditionally on μ), Y has a Poisson distribution, but the mean itself varies according to a gamma distribution with shape parameter k. Note that a negative binomial random variable can assume any possible non-negative integer value.

Remark 1. For any positive real number $c > 0$, the gamma function is $\Gamma(a) = \int_0^\infty x^{c-1} \exp(-x) dx$. Here are some properties of the Gamma function
(i) $\Gamma(c+1) = c\Gamma(c)$.
(ii) if c is a positive integer, then $\Gamma(c) = (c-1)!$.

For the negative binomial distribution, we have the following result.

Fact 2. If Y has a negative binomial distribution with parameters μ and k. Then

$$E(Y) = \mu, \text{ and } Var(Y) = \mu + D\mu^2, \text{ where } D = 1/k.$$

Note that the variance is larger than the mean since $D > 0$. Also note that when $k = \infty$ (ie, $D = 0$), the variance equal the mean. As a matter of fact it can be shown that when $k \to \infty$, the pdf $f(y; k, \mu) \to \exp(-\mu)\mu^y / y!$, the pdf of the Poisson(μ) distribution.

Remark 2.
(i) If k is a positive integer, then we may rewrite the negative binomial pdf as

$$\begin{aligned} f(y; k, \mu) &= \binom{y+k-1}{k-1} \pi^k (1-\pi)^y, \text{ with} \\ \pi &= \frac{k}{\mu+k}. \text{ [Note: } 0 < \pi < 1.] \end{aligned}$$

Suppose we have a coin with faces S (success) and F (failure) and we toss this coin till we have k S's. Sometimes this is called an inverse sampling scheme. Let Y be the number of F's till we get k S's. Assume that the probability of sucess on any toss equals π. When $Y = y$, say ($y = 3$), the total number of tosses is then $y + k$, the $(y + k)^{th}$ toss must result in an S, and there are $k - 1$ S's in the first $y + k - 1$ tosses. The number of ways the first $k - 1$ S's out of $y + k - 1$ tosses can come up is $\binom{y+k-1}{k-1}$. Therefore

$$P(Y = y) = \binom{y + k - 1}{k - 1}\pi^k(1 - \pi)^y.$$

(ii) Why is the name 'negative binomial'? The reason is technical. It can be shown that for any $0 < u < 1$,

$$(1 - u)^{-k} = \sum_{y=0}^{\infty}\binom{y + k - 1}{k - 1}u^k.$$

The last expansion is called a negative binomial expansion. Taking $u = 1 - \pi$, we have

$$\begin{aligned}
\pi^{-k} &= (1 - (1 - \pi))^{-k} = \sum_{y=0}^{\infty}\binom{y + k - 1}{k - 1}(1 - \pi)^y, \text{ ie,} \\
1 &= \sum_{y=0}^{\infty}\binom{y + k - 1}{k - 1}\pi^k(1 - \pi)^y = \sum_{y=0}^{\infty} f(y; k, \pi).
\end{aligned}$$

Thus the name comes from the negative binomial expansion since coefficients in the negative binomial expansion of $(1-(1-\pi))^{-k}$ are proportional to the pdf $f(\cdot; k, \pi)$. Incidentally, 'binomial(n, π)' distribution get its name from the binomial expansion

$$1 = (\pi + (1 - \pi))^n = \sum_{y=0}^{n}\binom{n}{y}\pi^k(1 - \pi)^{n-y}.$$

We will conclude this handout with an example of a data set where the negative binomial distribution is more appropriate than the Poisson distribution.

Example 2. (Frequency of knowing homicide victim) The table below summarizes responses of 1308 subjects to the question: Within the last 12 months, how many people have you known personally that were victims of homicide. So the variable Y is the number of murder victims (in the past year) known to a person. The table below shows responses by race, for those who identified their race as white or as black. The sample mean for the blacks was 0.522, with a variance of 1.150. The sample mean for the whites was 0.092, with a variance of 0.155. Note that in both cases, the variance is nearly twice as large as the mean, indicating the possibility that a Poisson modeling may not be appropriate. We will examine this data later in more detail later in the course whereby a comparison will be made between the races. But here we will only look at the data for blacks, and compare Poisson and negative binomial fits.

Response	Black	White
0	119	1070
1	16	60
2	12	14
3	7	4
4	3	0
5	2	0
6	0	1

The trotal number of blacsks in the sample was 159. If we use a Poisson model, then there is only one parameter and its estimate is $\hat{\mu} = 0.522$. If we use a negative binomial model, then the MLE for μ and k are $\hat{\mu} = 0.522, \hat{D} = 4.94$, resulting in a variance estimate of

$$\hat{\mu} + \hat{D}\hat{\mu}^2 = 0.522 + (4.94)(0.522)^2 = 1.87.$$

The following table lists the probabilities of the fitted Poisson and negative binomial distributions as well as expected counts. Clearly, the negative binomial is a much better fit for this data than Poisson.

Response	Count	Sample Proportion	Poisson Prob	NB Prob	Poisson Count	NB Count
0	119	0.748	0.593	0.772	94.3	122.8
1	16	0.101	0.309	0.113	49.2	17.9
2	12	0.076	0.081	0.049	12.9	7.8
3	7	0.044	0.014	0.026	2.2	4.1
4	3	0.019	0.002	0.015	0.3	2.4
5	2	0.013	0	0.009	0	1.4
6	0	0	0	0.006	0	0.9