# Common Discrete Distributions

Statistics 104

Autumn 2004

Taken from Statistics 110 Lecture Notes

# Common Discrete Distributions

There are a wide range of popular discrete distributions used in probability modelling and statistics. A subset which will be discussed here are

- Bernoulli and Binomial
- Geometric and Negative Binomial
- Poisson
- Hypergeometric
- Discrete Uniform

# Bernoulli and Binomial

**Bernoulli**: This distribution is useful for describing the results of a single trial that is either a success (Prob $= p$) or a failure (Prob $= 1 - p = q$).

$$p(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{Otherwise} \end{cases}$$

This is the model I've been using for flipping a biased coin. Its also the distribution of an indicator RV. (Denoted by $Ber(p)$)

$$E[X] = p; \quad Var(X) = p(1-p); \quad SD(X) = \sqrt{p(1-p)}$$

**Binomial**: Let $Y_1, Y_2, \ldots, Y_n$ be independent $Ber(p)$ RVs. Then

$$X = \sum_{i=1}^{n} Y_i$$

is a binomial RV (Denoted $Bin(n, p)$). Note that $Ber(p) = Bin(1, p)$

$X$ is the number of successes in $n$ independent, identical (same $p$) trials.

It is used in a wide range of problems:

- Medical: number of patients, out of 500 in a trial, surviving 5 years
- Quality control: sample 10 parts and record the number that are faulty
- Coin flipping: number of heads in 50 flips.

Note in this second case, what we are calling a success (what we are counting) is actually a failure for a practical purposes.

The PMF for the binomial distribution is

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}; \quad k = 0, 1, \ldots, n$$

Table A in the back of the book gives the CDF for the binomial distribution for some values of $n$ (5, 10, 15, 20,2 5) and $p$ (0.01, 0.05, 0.10, ... , 0.9, 0.95, 1.0)

The following is copy of the table for $n = 5$

| $k$ \ $p$ | 0.01 | 0.05 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.951 | 0.774 | 0.590 | 0.328 | 0.168 | 0.078 | 0.031 | 0.010 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.999 | 0.977 | 0.919 | 0.737 | 0.528 | 0.337 | 0.188 | 0.087 | 0.031 | 0.007 | 0.000 | 0.000 | 0.000 |
| 2 | 1.000 | 0.999 | 0.991 | 0.942 | 0.837 | 0.683 | 0.500 | 0.317 | 0.163 | 0.058 | 0.009 | 0.001 | 0.000 |
| 3 | 1.000 | 1.000 | 1.000 | 0.993 | 0.969 | 0.913 | 0.812 | 0.663 | 0.472 | 0.263 | 0.081 | 0.023 | 0.001 |
| 4 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.990 | 0.969 | 0.922 | 0.832 | 0.672 | 0.410 | 0.226 | 0.049 |

Note that the row for $k = n$ must be all 1, so it is omitted.

| $k$ \ $p$ | 0.01 | 0.05 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.951 | 0.774 | 0.590 | 0.328 | 0.168 | 0.078 | 0.031 | 0.010 | 0.002 |
| 1 | 0.999 | 0.977 | 0.919 | 0.737 | 0.528 | 0.337 | 0.188 | 0.087 | 0.031 |
| 2 | 1.000 | 0.999 | 0.991 | 0.942 | 0.837 | 0.683 | 0.500 | 0.317 | 0.163 |
| 3 | 1.000 | 1.000 | 1.000 | 0.993 | 0.969 | 0.913 | 0.812 | 0.663 | 0.472 |
| 4 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.990 | 0.969 | 0.922 | 0.832 |

The table can be used to get the following probabilities (for $n = 5$ and $p = 0.6$)

$$
\begin{aligned}
P[X \leq 3] &= 0.663 \\
P[X = 2] &= P[X \leq 2] - P[X \leq 1] = 0.317 - 0.087 = 0.230 \\
P[X > 0] &= 1 - P[X \leq 0] = 1 - 0.010 = 0.90
\end{aligned}
$$

Statistics packages are better than tables, as they don't limit the choice of $n$ and $p$.

The moments of the distribution are

$$E[X] = np; \quad Var(X) = np(1-p); \quad SD(X) = \sqrt{np(1-p)}$$

The value for $E[X]$ can be derived a couple of ways. One is by the definition

$$\begin{aligned}
E[X] &= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= n \sum_{i=1}^{n} \binom{n-1}{i} pp^{i-1}(1-p)^{(n-1)-(i-1)} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} \\
&= np \sum_{j=0}^{n-1} P[Y=j] \qquad \text{where } Y \sim Bin(n-1,p) \\
&= np
\end{aligned}$$

or we can use the fact that

$$X = \sum_{i=1}^{n} Y_i$$

and

$$E[X] = \sum_{i=1}^{n} E[Y_i]$$

As mentioned earlier, $E[Y_i] = p$ and there are $n$ of them in the sum.

The variance can be derived a couple of ways as well. One approach uses the fact that

$$E[X^2] = np + n(n-1)p^2$$

which can be determined a number of ways. From this

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 = np + n(n-1)p^2 - (np)^2 \\ &= np - np^2 = np(1-p) \end{aligned}$$

Another approaches uses the fact that if $Y_1, Y_2, \ldots, Y_n$ are independent RVs and $X = \sum_i Y_i$

$$Var(X) = \sum_{i=1}^{n} Var(Y_i)$$

A $Ber(p)$ has variance $p(1-p)$, thus this approach gives us the same answer.

Note that it's variances that add, not standard deviations.

$$SD(X+Y) \neq SD(X) + SD(Y)$$

# Geometric and Negative Binomial Distribution

**Geometric**: This distribution is also based on Bernoulli trials. In this case, we repeat the trials until we get a success and record which trial we get this first success.

The flip a coin until the first head example used earlier is an example of the geometric distribution

The PMF for the geometric distribution ($Geo(p)$) is

$$p(k) = p(1-p)^{k-1}; \quad k = 1, 2, \ldots$$

The moments of the distribution are

$$E[X] = \frac{1}{p}; \quad Var(X) = \frac{1-p}{p^2} = \frac{1}{p}\left(\frac{1}{p} - 1\right); \quad SD(X) = \frac{\sqrt{1-p}}{p}$$

The mean suggests that if the chance of a success is 1 in $n$, on average we need to wait $n$ trials until we see our first success.

For example, if we bought one ticket a week in the Mass Millions lottery ($p = \frac{1}{13{,}983{,}816}$) we expect to wait 13,983,816 weeks (268,919 years) before we win the grand prize for the first time.

For the smallest prize (3 out of 6), $p = 0.0177$, so we expect to wait $\frac{1}{0.0177} = 56.66$ weeks until our first win of this type. As the standard deviation for this case is 55.99, the actual time we need to wait could be much higher.

There is a probability of 0.027 that the first 3 out of 6 win will take over 200 weeks.

**Negative Binomial**: This is similar to the geometric distribution except it is the time for the $r$th success.

The PMF for the negative binomial distribution ($NBin(r, p)$) is

$$p(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}; \quad k = r, r+1, r+2, \ldots$$

Note that $Geo(p) = NBin(1, p)$. Also $NBin(r, p)$ can be thought of as the sum of $r$ independent $Geo(p)$ RVs.

Note that some books and software define the geometric and negative binomial slightly differently. Instead of the waiting time until success $r$, they define it as the number of failures before success $r$. For example, the package `R` uses this alternate definition for its routines.

The moments of the distribution are

$$E[X] = \frac{r}{p}; \quad Var(X) = \frac{r(1-p)}{p^2} = \frac{r}{p}\left(\frac{1}{p} - 1\right); \quad SD(X) = \frac{\sqrt{r(1-p)}}{p}$$

These results come from the fact that the negative binomial is the sum of $r$ geometric RVs.

# Poisson

This is a distribution that is used for counts, often of rare events. It was originally derived as a limiting distribution of the binomial by Poisson (see the argument in section 2.1.5). However it is useful for many situations. One example mentioned before was the number of alpha particles emitted from a radioactive source during a fixed time period.

Other examples:

- Queuing theory. How many customers will enter a bank in the next 30 minutes.
- Wave damage to cargo ships
- Number of matings per year in a population of African elephants

The PMF for the Poisson distribution ($Pois(\lambda)$) is

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}; \quad k = 0, 1, 2, \ldots$$

The parameter of the Poisson distribution is the mean.

Note that often when the Poisson distribution is used, the response is the number of counts during some time period. So lambda is often given as

$$\lambda = rate \times time$$

So doubling the time period observed will double $\lambda$

The moments of the distribution are

$$E[X] = \lambda; \quad Var(X) = \lambda; \quad SD(X) = \sqrt{\lambda}$$

$$\begin{aligned} E[X] &= \textstyle\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} \\ &= \lambda \textstyle\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} \\ &= \lambda \textstyle\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}e^{-\lambda} = \lambda \end{aligned}$$

$$\begin{aligned} E[X(X-1)] &= \textstyle\sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} \\ &= \lambda^2 \textstyle\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} \\ &= \lambda^2 \end{aligned}$$

$$E[X^2] = E[X(X-1)] + E[X] = \lambda^2 + \lambda$$

which gives

$$Var(X) = E[X^2] - (E[X])^2 = \lambda$$

Example: Bomb hits in London during WWII (R.D. Clark, 1946)

London was divided into $N = 576$ small blocks of $0.25(km)^2$

Let $\lambda$ be the average number of hits in a block by bombs. The estimate of this is $\hat{\lambda} = 0.9323$.

For any such block, the $X$ be the number of hits. $X \sim Pois(\lambda)$ (approximately).

Compare with actual data ($N_k = \#$ blocks with $k$ hits)

| $k$ | $N_k$ | $N \times P[X = k]$ |
|---|---|---|
| 0 | 229 | 226.74 |
| 1 | 211 | 211.39 |
| 2 | 93 | 98.54 |
| 3 | 35 | 30.62 |
| 4 | 7 | 7.14 |
| 5+ | 1 | 1.57 |

# Hypergeometric

The binomial distribution occurs when you sample from a population consisting of two types of objects ("Success" & "Failure") **with replacement**. Most sampling however is done **without replacement**. The **Hypergeometric** distribution describes this situation.

Assume that your population contains $n$ members, $r$ of which are "Successes" and $n - r$ of them are "Failures". You want to draw $m$ items from the population. Let $X$ be the number of successes in the $m$ draws.

The PMF for the hypergeometric distribution is

$$p(k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}; \quad k = 0, 1, 2, \ldots, m$$

An example of the hypergeometric distribution was the Mass Millions lottery. The probability of the the number of matches is hypergeometric with $n = 49$, $r = 6$, and $m = 6$

| $k$ | $P[X = k]$ |
|---|---|
| 0 | 0.43596 |
| 1 | 0.41302 |
| 2 | 0.13238 |
| 3 | 0.01765 |
| 4 | 0.00097 |
| 5 | 0.000018 |
| 6 | 1/13,983,816 |

When $n$, $r$, and $n - r$ get big relative to $m$, the hypergeometric looks a lot like the $Bin(m, p = \frac{r}{n})$ distribution. In this situation, the probability of a success doesn't change much as members of the population are sampled. For the above example however the binomial approximation is poor.