Poisson, Negative Binomial and Hypergeometric

Math 477

November 26, 2014

1 The Poisson

Definition 1.1. *X is a Poisson RV with parameter* $\lambda, \lambda > 0$*, denoted as Poisson(*λ*) if*

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2\cdots$$

To check that the above defines a distribution, we need the following Lemma

Lemma 1.2.

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}.$$

Proof. Define

$$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

We see that $f(0) = 1$. Moreover,

$$\begin{aligned} \frac{d}{dx}f(x) &= \sum_{k=0}^{\infty} \frac{d}{dx}\frac{x^k}{k!} \\ &= \sum_{k=1}^{\infty} \frac{d}{dx}\frac{x^{k-1}}{(k-1)!} \\ &= \sum_{k=0}^{\infty} \frac{d}{dx}\frac{x^k}{k!} = f(x). \end{aligned}$$

By uniqueness of solution to an initial value ODE problem, we conclude $f(x) = e^x$.

1.1 Expectation

Lemma 1.3. *Let X be a Poisson(λ) RV. Then $E(X) = \lambda$.*

Proof.

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

1.2 Variance

Lemma 1.4. *Let X be a Poisson(λ) RV. Then $Var(X) = \lambda$.*

Proof.

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

Thus

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

1.3 Poisson as approximation to Binomial distribution

1.3.1 The approximation

Let X be a Binomial(n,p) RV such that $np = \lambda$, n is large and p is small (think $n \to \infty$ and $p := \lambda/n$.) Then

$$\begin{aligned} P(X=k) &= \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\frac{\lambda^k}{k!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}. \end{aligned}$$

As $n \to \infty$, $\frac{\left(1-\frac{\lambda}{n}\right)^n}{\to} e^{-\lambda}$. Thus $P(X=k) \approx e^{-\lambda}\frac{\lambda^k}{k!}$.

1.3.2 Some examples

The following scenarios may be modeled as a Poisson RV (or rather a Possion approximation to a Binomial)

- The number of misprints on a page (or a group of pages) of a book
- The number of people in a community who survive to 100
- The number of wrong telephone numbers that are dialed in a day
- The number of packages sold in a particular store each day
- The number of customers entering a post office on a given day
- The number of α-particles discharged in a fixed period of time from some radioactive material
- The number of officers died by horsekick in a battalion.

In short, the Poisson RV is used to model rare events that happens with some given average (λ).

Example 1.5. *The average number of typo on a page of a math book is 1/2. What is the probability that there are at least 1 error on the first page of your textbook?*

Ans: Let's just explain again why Poisson(1/2) is the appropriate distribution to use here. There are many (n) words on a particular page of a textbook, and the probability (p) of any word being a typo is small, but the expectation (i.e. the average number of typos) is $np = 1/2$.

We have $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/2} \approx .393$.

Example 1.6. *The average number of customers arriving at lunch time at the New Brunswick post office is 1. What is the probability that there is exactly 1 customer arriving at New Brunswick post office at lunch time today?*

Ans: Poisson(1) is the appropriate distribution to use. Imagine there can be many (n) custommers (potentially those living in New Brunswick) arriving at the Post office during lunch (potentially ∞!). The probability that any customer going to the Post office at lunch is p, and this p is such that $np = 1$.

We have $P(X = 1) = e^{-1}$.

Remark: You may wonder how to estimate the actual λ. Here's a possible way: for the post office example record the number of customers arriving at the Post Office during lunch during a 30 days interval. Take that average, it should be approximately λ.

For the typo example, count the number of typos in 30 pages or so. Take the average, again it should be approximately λ.

These are examples of the Law of Large Number: the average over time is approximately equal to the expectation of the random variable. We'll cover it later on in this semester.

Example 1.7. *The average number of officers died by horse kick at a particular batallion is 3. What is the probability that 2 officers will die by horse kick this year?*

Ans: Imagine there are many (n) officers in the batallion, and each has a small probability (p) of being dead by horse kick, such that $np = 3$.

Then $P(X = 2) = e^{-3}\frac{3^2}{2!}$.

1.4 Poisson as model of arrival of events

We have mentioned the example of using Poisson random variable to model Post office customers arrival at lunch time. We argued it using the Binomial approximation. It turns out that the Poisson random variable can be used in a very similar way to model events arriving at a certain rate, for example earthquake, people arriving at a particular establishment etc.

1.5 Description of the model

Let us fix a $\lambda > 0$ to be the rate of the event arrival (λ has the unit of something / time, for example 1 person / min, 2 earthquakes / year etc.) We make the following assumptions:

- For small h, the probability that exactly 1 event occurs in a given interval of length h is $\lambda h + o(h)$, where $o(h)$ denotes quantity very small with respec to h ($o(h)/h \to 0$ as $h \to 0$, like h^2).
- The probability that 2 or more events occur in an interval of length h is equal to $o(h)$.
- The number of events happening in non-overlapping intervals are independent.

  Now let us fix a time s. We would like to find the distribution of the number of events happening between s, t, denoted as $N(s, t)$. We have

  Lemma 1.8. *If the above assumptions are satisfied, then $N(s, t)$ has distribution Poisson$(\lambda(t - s))$.*

  Proof. We divide the interval (s, t) into n subintervals, each with length $(t - s)/n$. For a fixed k, we want to compute $P(N(s, t) = k)$. Let E be the event that at most one event occurs within each interval. Then

  $$P(N(s,t) = k) = P(N(s,t) = k, E) + P(N(s,t) = k, E^c).$$

  We show $P(E^c) = 0$, as $n \to \infty$. Note that

  $$\begin{aligned} P(E^c) &\leq nP(\text{ more than 1 events occur in the first interval}) \\ &\leq n\, o((t-s)/n) \to 0 \text{ as } n \to \infty. \end{aligned}$$

  Now

  $$P(N(t,s) = k, E) = \binom{n}{k} \left(\frac{\lambda(t-s)}{n} + o((t-s)/n) \right)^k \left(1 - \frac{\lambda(t-s)}{n} + o((t-s)/n) \right)^{n-k}.$$

  Denoting $p = \frac{\lambda(t-s)}{n} + o((t-s)/n)$ we see that $pn \to \lambda$ as $n \to \infty$. Thus arguing almost the same as the Binomial approximation,we see that

  $$P(N(t,s) = k) \approx e^{-\lambda} \lambda^k k!.$$

Example 1.9. *Earth quake occurs in California with rate 2 per week. What is the probability that there will be no earth quake during the next month?*

Ans: The number earth quake within next month is Possion (8). Thus

$$P(X = 0) = e^{-8}.$$

Remark: Let T be the number of weeks until the next earthquake (T is a random variable). If $T > t$ it means that there is no earthquake within t weeks. That is $P(T > t) = e^{-2t}$. We also have $P(T \le t) = 1 - e^{-\lambda t}$.

2 The negative binomial

Definition 2.1. *X is a Negative Binomial with parameters $p, r, 0 \le p \le 1, r$ integer, $r \ge 1$ denoted as Negative Bin(r, p) if*

$$\begin{aligned} P(X = k) &= \binom{k-1}{r-1}(1-p)^{k-r}p^{r-1}p \\ &= \binom{k-1}{r-1}(1-p)^{k-r}p^{r}, k = r, r+1, r+2, \cdots \end{aligned}$$

Remark: The negative binomial represents the number of trials until the first r successes, if the success probability of each trial is p.

We can represent a Negative Binomial (r,p) as followed:

Proposition 2.2. *Let $Y_1, Y_2, \cdots, Y_r$ be independent Geometric(p). Then $X = \sum_{i=1}^{r} Y_i$ has Negative Bin(r,p) distribution.*

Proof.

$$\begin{aligned} P(\sum_{i=1}^{r} Y_i = k) &= \sum_{k_1, k_2, \cdots k_r : \sum_i k_i = k} P(Y_i = k_i, i = 1, \cdots r) \\ &= \sum_{k_1, k_2, \cdots k_r : \sum_i k_i = k} \prod_{i=1}^{r} (1-p)^{k_i - 1} p \\ &= \sum_{k_1, k_2, \cdots k_r : \sum_i k_i = k} \prod_{i=1}^{r} (1-p)^{k_i - 1} p \\ &= \sum_{k_1, k_2, \cdots k_r : \sum_i k_i = k} (1-p)^{k-r} p^r \\ &= \binom{k-1}{r-1}(1-p)^{k-r}p^r, \end{aligned}$$

where the last inequality is because the number of ways we can write a sum of r positive integers adding up to k is exactly $\binom{k-1}{r-1}$.

Remark 2.3. *For $k \geq r$, let Y have Bin(k,p) distribution. Then*

$$P(Y = r) = \binom{k}{r} p^r (1-p)^{k-r}.$$

On the other hand, the probability that r successes happen at the first $k-1$ trials is

$$\binom{k-1}{r} p^r (1-p)^{k-r}.$$

Substracting these two probabilities gives the probability that we have 5 successes, but the last one is on the k trial:

$$\binom{k}{r} p^r (1-p)^{k-r} - \binom{k-1}{r} p^r (1-p)^{k-r} = \binom{k-1}{r-1} p^r (1-p)^{k-r},$$

by the identity

$$\binom{k-1}{r-1} + \binom{k-1}{r} = \binom{k}{r}.$$

But this is also the probability that a Neg Bin (r, p) takes value k.

Remark 2.4. *Let $Y = X - r$ be the number of failures before r successes. Then it is clear that*

$$\begin{aligned} P(Y = k) = P(X = k + r) &= \binom{k+r-1}{r-1} p^r (1-p)^{k-r} \\ &= \binom{k+r-1}{k} p^r (1-p)^{k-r} \end{aligned}$$

There is a well-known identity in combinatorics that says

$$\binom{-r}{k} = (-1)^k \binom{k+r-1}{k} = \left(\!\binom{r}{k}\!\right),$$

where the last notation denotes a multiset coefficient: it counts the number of combinations you can get from a sample of size k from r objects, with replacement (so with repetition in your sample). For example, from the set $\{1, 2\}$, if we pick 3 objects with replacement then all possible outcomes would be $\{1, 1, 2\}, \{2, 2, 1\}, , \{1, 1, 1\}, \{2, 2, 2\}$.

In general, the number of ways to pick k objects out of r objects with replacement is the same as finding the number of ways to sum $\sum_{i=1}^{r} a_i = k$ where $a_i \geq 0$. The *interpretation is a_i represents the number of the ith object in the sample. We have argued before that this number of way is $\binom{k+r-1}{r-1} = \binom{k+r-1}{k}$.*

Finally, the name negative binomial comes from the relation

$$\binom{-r}{k} = (-1)^k \binom{k+r-1}{k}.$$

2.1 Expectation

Lemma 2.5. *Let X be a Negative Bin(r,p) RV. Then $E(X) = \frac{r}{p}$.*

Heuristically we can argue it as followed:

$$E(X) = E(Y_1 + Y_2 + \cdots + Y_r) = 1/p + 1/p + \cdots 1/p = r/p.$$

For a rigorous proof, you can see the textbook, which uses a similar technique as when we computed the Binomial.

2.2 Variance

Lemma 2.6. *Let X be a Negative Bin(r,p) RV. Then $Var(X) = r(\frac{1}{p} - \frac{1}{p^2})$.*

Again, heuristically, since Y_i are independent:

$$Var(X) = Var(Y_1 + Y_2 + \cdots + Y_r) = (\frac{1}{p} - \frac{1}{p^2}) + (\frac{1}{p} - \frac{1}{p^2}) + \cdots (\frac{1}{p} - \frac{1}{p^2}) = r(\frac{1}{p} - \frac{1}{p^2}).$$

For a rigorous proof, you can also see the textbook.

2.3 Example

Example 2.7. *At all times, a pipe-somking professor carries 2 match boxes - 1 in his left hand pocket and one in his right hand pocket. Each time he needs a match, he is equally likely to take it from either pocket. Suppose the mathematician first discovers that one of his match boxes is empty and both boxes contained N matches initially. What is the probability that there are exactly k matches, $k = 0, 1, \cdots, N$ in the other box?*

Ans: Let E be the event that the right hand box is empty (then E^c is the event that the left hand box is empty, by hypothesis). We have

P(k matches remaining) = P(k matches remaining $|E$) P(E) + P(k matches remaining $|E^c$) P(E^c).

By symmetry of the problem, P(k matches remaining $|E$) = P(k matches remaining $|E^c$) and $P(E) = P(E^c) = 1/2$. Thus

P(k matches remaining) = P(k matches remaining $|E$).

Now P(k matches remaining $|E$) = P(k matches remaining, $E)/P(E)$ = 2 P(k matches remaining, E).

If we have k matches remaining in the left hand pocket and the right hand box is empty, out of $N + (N - k) + 1$ trials, (we must add the last trial where he discovers that the box is empty) we must have picked N from the right, $N - k$ from the left and the last one from the right (the one that he discovers it is empty).

If we think of picking from the right as "success" then this is the setting of the negative geometric. Thus the probability of interest is

$$2\binom{2N+1-k}{N}(1/2)^{N+1}(1/2)^{N-k} = \binom{2N+1-k}{N}(1/2)^{2N-K}.$$

2.4 The hypergeometric

Definition 2.8. *X is a Hypergeometric RV with parameters N, m, n; N, m, n integers, $0 \le m, n \le N$ if*

$$P(X = k) \quad = \quad \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, k = 0, \cdots, n.$$

Remark: 1. Consider the experiment of picking n balls out of N balls, out of which m are white, $N - m$ are black without replacement. The Hypergeometric represents the number of white balls we have in our sample.

2. It could be that $m < n$, that is the number of white balls available in the urn is less than the sample size. In this case by convention we say $\binom{n}{k} = 0$ if $k > n$, which implies $P(X = k) = 0$ for $k > m$.

Again, it is not easy to establish analytically that

$$\sum_{k=0}^{n} \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = 1,$$

but heuristically, you can see that the number of ways we can pick n objects out of N objects, which consist of 2 groups, equal to the number of ways we can pick the first k objects out of the first group and the first $n - k$ objects out of the 2nd group. That is

$$\sum_{k=0}^{n} \binom{m}{k}\binom{N-m}{n-k} = \binom{N}{n}.$$

2.5 Expectation

Lemma 2.9. *Let X be a Hypergeometric(N,m,n) RV. Then $E(X) = n\frac{m}{N}$.*

Proof.

$$E(X) = \sum_{k=0}^{n} k\binom{m}{k}\binom{N-m}{n-k} \Big/ \binom{N}{n}$$

Using the identity

$$k\binom{m}{k} = m\binom{m-1}{k-1},$$

we have

$$E(X) = \sum_{k=0}^{n} m\binom{m-1}{k-1}\binom{(N-1)-(m-1)}{(n-1)-(k-1)} \Big/ \frac{N}{n}\binom{N-1}{n-1} = n\frac{m}{N}$$

2.6 Variance

Lemma 2.10. *Let X be a Hypergemotric(N,m,n) RV. Then*

$$Var(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{(N-1)} + 1 - \frac{nm}{N}\right].$$

If we denote $p = m/N$ and observe that

$$\frac{m-1}{N-1} = \frac{Np-1}{N-1} = p - \frac{1-p}{N-1},$$

then we also have

$$Var(X) = np(1-p)\left(1 - \frac{n-1}{N-1}\right).$$

Proof: See textbook. It relies on showing that

$$E(X^k) = \frac{nm}{N}E((Y+1)^{k-1}),$$

where Y is a Hypergeometric (N-1, m-1, n-1).

2.7 Binomial approximation to Hypergeometric

The expression $E(X) = np$ and

$$Var(X) = np(1-p)\left(1 - \frac{n-1}{N-1}\right)$$

suggests that the Hypergeometric is close to a Binomial(n, m/N) if N is large relative to n, keeping p constant. Indeed this is the case when we sample a large population. We sample without replacement, but we treat the resulting distribution as a Binomial(n,p). The calculation for the approximation is as followed:

$$\begin{aligned}
P(X=k) &= \binom{m}{k}\binom{N-m}{n-k}\Big/\binom{N}{n} \\
&= \frac{m!}{(m-k)!k!}\frac{(N-m)!}{(N-m-n+k)!(n-k)!}\frac{(N-n)!n!}{N!} \\
&= \frac{n!}{(n-k)!k!} \times \frac{m}{N}\frac{m-1}{N-1}\cdots\frac{m-k+1}{N-k+1} \\
&\quad \times \frac{N-m}{N-k}\frac{N-m-1}{N-k-1}\cdots\frac{N-m-(n-k-1)}{N-k-(n-k-1)}
\end{aligned}$$

$$\approx \binom{n}{k}p^k(1-p)^{n-k}.$$