Random Variables Lecture 8

The Poisson Random Variable

A random variable X that takes on one of the values $0, 1, 2, \dots$ is said to be a Poisson random variable with parameter λ if $p(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$

Clearly $\sum_{k=0}^{\infty} P(X=k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$

so $p: \mathbb{N} \to \mathbb{R}$ $\quad p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ is indeed a probability mass function.

The Poisson Random Variable is very useful because it serves as an approximation to binomial and nearly binomial distributions, when n - the number of trials - is large and p, the probability of success for each individual trial is small so that $\lambda = np \ll n$.

<u>Thm:</u> Let X be binomial r.v. with parameters (n,p) if $np \ll n$ then $P(X=k) \approx e^{-\lambda}\frac{\lambda^k}{k!}$ where $\lambda = np$ and $k \ll n$.

<u>Proof:</u> $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$

$$= \frac{n(n-1)\dots(n-k+1)}{k!}\left(\frac{np}{n}\right)^k \left(1-\frac{np}{n}\right)^{n-k}$$

$$= \frac{n(n-1)\dots(n-k+1)}{n^k} \cdot \lambda^k \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \cdot \frac{1}{k!}$$

$$= \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-k+1}{n} \cdot \lambda^k \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \cdot \frac{1}{k!}$$

Notice that $\lim_{n\to\infty} \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-k+1}{n} = 1$

and $\lim_{n\to\infty} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} = \frac{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n}{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^k} = \frac{e^{-\lambda}}{1}$

Thus for $n \gg \lambda$ and $n \gg k$ we have

$$\frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-k+1}{n} \lambda^k \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \cdot \frac{1}{k!} \approx e^{-\lambda} \frac{\lambda^k}{k!}$$

Ex. A newly published book is carefully examined and found to contain on average $\frac{1}{2}$ typographical errors per page. You open to page 100, what is the probability that you will find one or more typographical errors on this page?

Solution: This page contains a large number of letters. Let's say this number is n. The probability that a letter is out of place is p and it is small. Now, we do not know n or p, but the average, np, is $\frac{1}{2}$. So we can use Poisson! let X be Poisson r.v. with parameter $\lambda = np = \frac{1}{2}$.

Then $P(X \geq 1) = 1 - P(X=0) = 1 - e^{-\frac{1}{2}} \approx 0.393$

Thus the probability of finding at least one typo is approximately 0.39.

<u>Ex.</u> Consider an experiment that consists of counting the number of α particles given off in a 1-sec interval by 1 gram of radioactive material. If we know from past experience that, on average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than 2 α particles will appear?

<u>Solution:</u> 1 gram of matter contains huge many atoms ① ② ... ⓝ. Each atom is like a trial. It either releases an α particle (success) or fails to release an α particle (failure). The probability of success is some unknown p, and the setup looks very much like that of a binomial distribution. From experiment the expected number of successes is $np = 3.2$.

Let X be Poisson with parameter $\lambda = 3.2$. We wish to compute $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$

$$= e^{-3.2} + e^{-3.2} \cdot 3.2 + e^{-3.2}\frac{(3.2)^2}{2} = e^{-3.2}\left(1 + 3.2 + \frac{(3.2)^2}{2}\right)$$

≈ 0.3799.

Ex. Every day a million people decide whether to visit a particular website. Assuming that each person makes the dicision to log on independently with probability $p = 2\times 10^{-6}$, estimate the probability of at least 3 visits.

Solution: The actual distribution of visits is binomial with 10^6 trials and success probability $p = 2\times10^{-6}$ for each trial. Thus the expected number of visits is $10^6 \cdot 2\times 10^{-6} = 2$.

The exact probability of 3 or more visits is

$$1 - \binom{10^6}{0}(1-p)^{10^6} - \binom{10^6}{1} p(1-p)^{10^6-1} - \binom{10^6}{2} p^2 (1-p)^{10^6-2}.$$

These calculations are hard to work with! For instance

$$(1-p)^{10^6} = \left(\frac{10^6-2}{10^6}\right)^{10^6}$$

requires raising large numbers to very heigh powers. On the other hand, if we let X be poisson with parameter $\lambda = 2$

$$P(X \leqslant 3) = 1 - P(X=0) - P(X=1) - P(X=2)$$

$$= 1 - e^{-2}\left(1 + 2 + \frac{2^2}{2}\right) = 1 - 5e^{-2} \approx 0.3233.$$

This is extremely accurate in light of the fact that

$2 = \lambda \ll 10^6 = n$ and $3 = k \ll 10^6 = n$.

The actual distribution doesn't have to be binomial to be accurately approximated by a Poisson distribution.

Ex. There are k distinguishable balls and n distinguishable boxes. The balls are randomly placed in the boxes, with all n^k possibilities equally likely. (Apparently problems of this type are at the core of many widely used algorithms in computer science).

(a) Find the expected number of empty boxes

(b) Find the probability that at least one box is empty.

(c) Let $n = 1000$ and $k = 5806$. Find a good approximation that at least one box is empty.

Solution:

(a) Let $X = \#$ of empty boxes. Then $X = X_1 + \ldots + X_n$ where $X_j = \begin{cases} 1 & \text{if } j^{th} \text{ box is empty} \\ 0 & \text{otherwise.} \end{cases}$

Then $E[X] = n E[X_1] = n\left(\frac{n-1}{n}\right)^k = n\left(1 - \frac{1}{n}\right)^k$

(b) Let E_j be the event j^{th} box is empty.

The desired probability is $p(E_1 \cup \ldots \cup E_n) =$

$$= \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<\dots<i_r} P(E_{i_1}\dots E_{i_r}) = \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} P(E_1 \dots E_r)$$

$$= \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \left(\frac{n-r}{n}\right)^k$$

(c) If $n = 1000$ and $k = 5806$ $E[X] = 1000\left(1 - \frac{1}{1000}\right)^{5806}$

$$= 1000\left[\left(1-\frac{1}{1000}\right)^{1000}\right]^{5806} = 1000\left(1-\frac{1}{1000}\right)^{5.806\cdot 1000}$$

$$\approx 1000\, e^{-5.806} \approx 3$$

Let Y be Poisson with parameter $\lambda = 3$. Then

$$P(X \geq 1) = 1 - P(X=0) \approx 1 - P(Y=0) = 1 - e^{-3} \approx 0.95.$$

<u>Ex.</u> Recall the birthday problem: If n people are in a room, what should be the minimal number of people to make it at least 50% likely that at least 2 people share a birthday?

<u>Solution:</u> If n people are present, there are $\binom{n}{2}$ birthday comparisons. The probability of an individual match is $\frac{365}{365^2} = \frac{1}{365}$ and the expected number of matches is $\binom{n}{2}\cdot\frac{1}{365} = \frac{n(n-1)}{2\cdot 365} = \lambda$. Let Y be Poisson with this parameter,

Then $P(Y \geq 1) \geq \frac{1}{2}$ means $1 - P(Y=0) \geq \frac{1}{2}$

or $\frac{1}{2} \geq e^{-\frac{n(n-1)}{2 \cdot 365}} \iff e^{-\ln 2} \geq e^{-\frac{n(n-1)}{2\cdot 365}}$

$\iff \frac{n(n-1)}{2 \cdot 365} \geq \ln 2$

$n(n-1) \geq 505.997$. $n^2 - n - 506 \geq 0$ or $n \geq \frac{1+\sqrt{1+4 \cdot 506}}{2}$

or $n \geq 23$.

Ex. What should be the number of people in the room, n, be so that the probability of them having the same or consecutive birthday is $\frac{1}{2}$ or more?

Solution: Just as before, there are $\binom{n}{2}$ trials, but this time the probability of success is $\frac{3 \cdot 365}{365^2} = \frac{3}{365}$ (why?) Working as before we get the inequality

$3n(n-1) \geq 506$ or $n \geq 14$.

Expectation and Variance of a Poisson Random Variable

We can anticipate both the expected value and Variance of a ~~binom~~ Poisson random variable with parameters λ by noting that $\lambda = np$ where (n, p) are the parameters of a binomial random variable X. If Y is Poisson with parameter λ, we anticipate $E[Y] = E[X] = np = \lambda$.

Since $Var(X) = np(1-p)$ where p is small, $Var(X) \approx np$ $= np = \lambda$. We therefore anticipate that $Var(Y) = \lambda$.

<u>Thm:</u> Let Y be Poisson with parameter λ. Then $E[Y] = Var(Y) = \lambda$

<u>Proof:</u> $$E[Y] = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

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

$$= e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

$$E[Y^2] = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!}$$

$$= \lambda(E[Y]+1) = \lambda(\lambda+1)$$

Thus $Var(Y) = E[Y^2] - (E[Y])^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$.

Properties of sum of Poisson Random Variables

<u>Thm:</u> If X is Poisson with parameter λ_1 and Y is Poisson with parameter λ_2 and these variables are independent, then $X+Y$ is Poisson with parameter $\lambda = \lambda_1 + \lambda_2$.

Proof: $$P(X+Y=k) = \sum_{j=0}^{k} P(X=j)P(Y=k-j)$$

$$= \sum_{j=0}^{k} \frac{e^{-\lambda_1}\lambda_1^j}{j!} \cdot \frac{e^{-\lambda_2}\lambda_2^{k-j}}{(k-j)!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} \lambda_1^j \lambda_2^{k-j}$$

$$= \frac{e^{-\lambda}}{k!} \sum_{j=0}^{k} \binom{k}{j} \lambda_1^j \lambda_2^{k-j}$$

$$= \frac{e^{-\lambda}}{k!} (\lambda_1+\lambda_2)^k = e^{-\lambda}\frac{\lambda^k}{k!}$$

Thm: If X and Y are Poisson r.v. with parameters λ_1 and λ_2 (X and Y are independent), then

$$P(X=k \mid X+Y=n) = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

is a binomial distribution.

Proof: $$P(X=k \mid X+Y=n) = \frac{P(X=k, X+Y=n)}{P(X+Y=n)}$$

$$= \frac{P(X=k, Y=n-k)}{P(X+Y=n)} = \frac{e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^k}{k!} \cdot \frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}}$$

$$= \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} \cdot \frac{1}{(\lambda_1+\lambda_2)^n} = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \cdot \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

Ex. Number of people entering Post office on any given day is a Poisson random variable with parameter λ. Show that if each person who enters the post office is a male with probability p and female with probability $1-p$, the number of males and females entering the office are independent Poisson random variables.

Solution: $X = \#$ males, $Y = \#$ females, $Z = \#$ people

Then $P(X=k) = P\left(\bigcup_{n=k}^{\infty} \{X=k, Z=n\}\right)$

$$= \sum_{n=k}^{\infty} P(X=k \mid Z=n) P(Z=n)$$

$$= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} = \sum_{n=k}^{\infty} e^{-\lambda p} \frac{(\lambda p)^k}{k!} \cdot$$

$$\cdot \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}.$$

Similarly, $P(Y=j) = e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!}$

To see that X and Y are independent, notice that

$$P(X=k, Y=j) = P(X=k, Y=j \mid Z=k+j) \cdot$$

$$\cdot P(Z=k+j) = \binom{k+j}{k} p^k (1-p)^j e^{-\lambda} \frac{\lambda^{k+j}}{(k+j)!}$$

$$= e^{-\lambda p} \frac{\lambda^k}{k!} e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!}$$