Lecture 8

The Geometric Distribution

The geometric distribution is a special case of negative binomial, it is the case $r=1$. It is so important we give it special treatment.

Motivating example

Suppose a couple decides to have children until they have a girl. Suppose the probability of having a girl is p. Let

X = the number of boys that precede the first girl

Find the probability distribution of X. First X could have any possible whole number value (although $X=1{,}000{,}000$ is very unlikely).

$$P(X=k) = P(\underbrace{\underline{B}\ \underline{B}\ \underline{B}\ \_\ \underline{B}}_{k}\ \underset{p}{\underline{G}})$$

$$= q^k p \quad (\text{where } q = 1-p)$$

We have suppose births are independent.

We have motivated

Definition

Suppose a discrete random variable X has the following pmf

$$P(X=k) = q^k p \,, \quad 0 \le k < \infty$$

The X is said to have geometric distribution with parameter p.

Remark

Usually this is developed by replacing "having a child" by a Bernoulli experiment and having a girl by a "success" (PC). I could have used coin flips.

Propositcon

Suppose X has geometric distribution with parameter p.

Then

(i) $E(X) = \dfrac{q}{p}$

(ii) $V(X) = \dfrac{q}{p^2}$

Proof of (i) (you are not responsible for this)

$$E(X) = (0)(p) + (1)(qp) + (2)(q^2p) + \cdots + (k)(q^kp) + \cdots$$
$$= p(q + 2q^2 + \cdots + kq^k + \cdots$$

Now $\dfrac{x}{(1-x)^2} = x + 2x^2 + 3x^3 + \cdots + kx^k + \cdots$ ← why?

so $E(X) = p\left(\dfrac{q}{(1-q)^2}\right) = p\left(\dfrac{q}{p^2}\right) = \dfrac{q}{p}$

$\square$

The Negative Binomial Distribution

Now suppose the couple decides they want more girls – say r girls; so they keep having children until the r-th girl appears. Let

$X =$ the number of boys that precede the r-th girl.

Find the probability distribution of X

Remark Sometimes (eg pg 13-14) it is better to write X_r instead of X.

Let's compute $P(X=k)$

$$\underbrace{\_\ \_\ \_\ \_\ \_\ \_\ \_}\ \underset{\substack{\uparrow \\ r\text{-th} \\ \text{girl}}}{\underline{G}}$$

What do we have preceding the r-th girl. Of course we must have $r-1$ girls and since we are assuming $X=k$ we have k boys so $k+r-1$ children.

All orderings of boys and girls have the same probability so

$$P(X=k) = (?)\,P(\underbrace{B \cdots B}_{k-1}\ \underbrace{G \cdots G}_{r-1}\ G)$$

or

$$P(X=k) = (?)\, q^{k} \cdot p^{r-1} \cdot q = (?)\, q^{k} p^{r}$$

$(?)$ is the number of words of length $k+r-1$ in B and G using k B's (whence $r-1$ G's).

Such a word is determined by choosing the slots occupied by the boys so there are $\binom{k+r-1}{k}$ words so

$$P(X=k) = \binom{k+r-1}{k} p^{r} q^{k}$$

So we have motivated the following

Definition

A discrete random variable X is said to have negative binomial distribution with parameters r and p if

$$P(X=k) = \binom{k+r-1}{k} p^r q^k, \quad 0 \leq k < \infty$$

The text denotes this pmf by $nb(x; r, p)$. so

$$nb(x; r, p) = \binom{x+r-1}{k} p^r q^x, \quad 0 \leq x < \infty.$$

Proposition

Suppose X has negative binomial distribution with parameters r and p. Then

(i) $E(X) = r\dfrac{q}{p}$

(ii) $V(X) = \dfrac{rq}{p^2}$

Waiting Times

The binomial, geometric and negative binomial distributions are all tied to repeating a given Bernoulli experiment (flipping a coin, having a child) infinitely many times.

Think of discrete time

$0, 1, 2, 3, \ldots$

and we repeat the experiment at each of these discrete times.

— E.g. flip a coin every minute

Now you can do the following things

1. Fix a time say n and let X = # of successes in that time period. Then $X \sim Bin(n,p)$. We should write X_n and think of the family of random variables parametrized by the discrete time n as the "binomial process". (see pg 18 — the Poisson process).

2. ((discrete) waiting time for the first success)

Let Y be the amount of time up to the time the first success occurs.

This is the geometric random variable. Why?

Suppose we have in our boy/girl example

$$\underbrace{\underset{0}{B}\quad \underset{1}{B}\quad \underset{2}{B}\quad \underset{3}{\_}\quad B}_{k}\quad \underset{k}{G}$$

So in this case

X = # of boys = k

Y = waiting time = k

so $$Y = X$$

3. Waiting time for r-th success

Now let Y_r = the waiting time up to the r-th success then there is a difference between X_r and Y_r

Suppose $X_r = k$

so there are k boys before the r-th girl arrives

$$\underbrace{\underset{0}{\_}\ \underset{1}{\underline{B}}\ \underset{2}{\_}\ \_\ \_\ \_\ \underset{k+r-2}{\_}}_{}\ \underset{k+r-1}{\underline{G}}$$

k B's r-1 G's so k+r-1 slots

But start at 0 so the last slot is

k+r - 2 so

$$Y_r = X_r + r - 1$$

The Poisson Distribution

For a change we won't start with a motivating example but will start with the definition.

Definition

A discrete random variable X is said to have Poisson distribution with parameter λ if

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad 0 \le k < \infty$$

We will abbreviate this to

$$X \sim P(\lambda).$$

I will now try to motivate the formula which looks complicated.

Why is the factor of $e^{-\lambda}$ there? It is there to make to total probability equal to 1.

$$\text{Total Probability} = \sum_{k=0}^{\infty} P(X=k)$$

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

But from calculus

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

$$\text{Total Probability} = e^{-\lambda} \cdot e^{\lambda} = 1$$

as it has to be.

Proposition

Suppose $X \sim P(\lambda)$. Then

(i) $E(X) = \lambda$

(ii) $V(X) = \lambda$

Remark It is remarkable that $E(X) = V(X)$.

Example 3.39

Let X denote the number of creatures of a particular type captured during a given time period. Suppose $X \sim P(4.5)$. Find $P(X=5)$ and $P(X \leq 5)$

Solution

$$P(X=5) = e^{-4.5}\frac{(4.5)^5}{5!}$$

(just plug into the formula using $\lambda = 4.5$)

$$P(X \le 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

$$= \underbrace{e^{-\lambda} + e^{-\lambda}\lambda + e^{-\lambda}\frac{\lambda^2}{2} + e^{-\lambda}\frac{\lambda^3}{3!} + e^{-\lambda}\frac{\lambda^4}{4!} + e^{-\lambda}\frac{\lambda^5}{5!}}_{\text{don't try to evaluate this.}}$$

The Poisson Process

A very important applicaton of the Poisson distributon arises in counting the number of occurrences of a certain event in time t

1. Animals in a trap
2. Calls coming into a telephone switchboard

Now we could let t vary so we get a one-parameter family of Poisson random variable

$$X_t \, , \quad 0 \le t < \infty \, .$$

Now a Poisson process is completely determined once we know its mean λ.

So for each t, X_t is a Poisson random variable. so

$X_t \sim P(\lambda(t))$

So the Poisson parameter λ is a function of t.

In the <u>Poisson process</u> ~~one~~ one assume that $\lambda(t)$ is the simplest possible function of t (aside from a constant function) namely a linear function

$$\lambda(t) = \alpha t.$$

Necessarily

$\alpha = \lambda(1) =$ the average number of observations in unit time.

Remark

In the text, page 124, the author proposes 3 axioms on a one parameter family of random variables X_t so that X_t is a Poisson process i.e

$$X_t \sim P(\alpha t)$$

Example (from an earlier version of the text)

The number of tickets issued by a meter reader can be modelled by a Poisson process with a rate of 10 ticket every two hours.

(a) What is the probability that exactly 10 tickets are given out during a particular 12 hour period.

Solution

We want $P(X_{12} = 10)$

First find $\alpha =$ average # of tickets by unit time

so $$\alpha = \frac{10}{2} = 5$$

so $X_t \sim P(5t)$

so $X_{12} \sim P((5)(12)) = P(60)$

$$P(X_{12}=10) = e^{-\lambda} \frac{\lambda^{10}}{(10)!}$$

$$= e^{-60} \frac{(60)^{10}}{(10)!}$$

(b) What is the probability that at least 10 tickets are given out during a 12 hour time period.

We want

$$P(X_{12} \geq 10) = 1 - P(X \leq 9)$$

$$= 1 - \sum_{k=0}^{9} e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= 1 - \underbrace{\sum_{k=0}^{9} e^{-60} \frac{(60)^k}{k!}}$$

not something you want to try to evaluate by hand.

Waiting Times

Again there are waiting time random variables associated to the Poisson process.

Let Y = waiting time until the first animal is caught in the trap

and Y_r = waiting time until the r-th animal is caught in the trap

Now Y and Y_r are continuous random variables which we are about to study. Y is exponential and Y_r has a special kind gamma distribution.