§2.1.3. The Geometric and Negative Binomial Distributions.

Consider independent Bernoulli trials in infinite sequence. On each trial, a success occurs with prob. p, and X is the total number of trials up to and including the first success.

$X=k$, there must be $k-1$ failures followed by a success.

$\{X=k\} = F_1 F_2 \cdots F_{k-1} S_k$, $\{X=1\}=S_1$.
$\{X=2\}=F_1 S_2$,
$\{X=3\}=F_1 F_2 S_3$.

$p(k) = P(X=k) = (1-p)^{k-1} \cdot p$, $k=1,2,3,\cdots$ called Geometric r.v.

verify $\sum_{k=1}^{\infty} p(k)=1$, $\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty}(1-p)^{k-1} \cdot p = p \cdot \sum_{k=1}^{\infty}(1-p)^{k-1}$

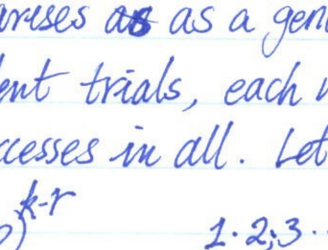
$$\overset{q=1-p}{=} p\cdot \sum_{k=1}^{\infty} q^{k-1} = p\,(1+q+q^2+\cdots) = p \cdot \frac{1}{1-q} = p \cdot \frac{1}{p} = 1.$$

Expl. A. The prob. of winning in a certain state lottery is said to be about $\frac{1}{9}$. If it is exactly $\frac{1}{9}$, the dist. of number of tickets a person must purchase up to and including the first winning ticket is a geometric r.v. with $p=\frac{1}{9}$.

$$p(k) = P(X=k) = \left(1-\frac{1}{9}\right)^{k-1} \cdot \frac{1}{9} = \left(\frac{8}{9}\right)^{k-1} \cdot \frac{1}{9}, \quad k=1,2,3,\cdots$$

use R: $p(k) = (1-p)^k \cdot p$. $k=0,1,2,\cdots$

```
k <- 0:50
pk <- dgeom(k, 1/9)
plot(k, pk, type="h")
```



or
```
k <- 1:50
pk <- (8/9)^(k-1) * 1/9
```

The negative binomial distribution arises ~~as~~ as a generalization of geometric dist. Suppose that a sequence of independent trials, each with prob. of success p, is performed until there are r successes in all. Let X be the total number of trials.

$$P(X=k) = \binom{k-1}{r-1} \underbrace{p^r \cdot (1-p)^{k-r}}_{\text{prob. for each sequence}}, \quad k=r, r+1, \cdots$$

$\underbrace{1 \cdot 2 \cdot 3 \cdots\cdots k-1}_{k-1 \text{ trials with } r-1 \text{ successes}} \cdot \overset{S}{k}$

Expl. B. Continuing Expl. A, the dist. of the number of tickets purchased up to and including the second winning ticket is negative binomial

$$p(k) = P(X=k) = \binom{k-1}{2-1} p^2 \cdot (1-p)^{k-2} = (k-1) \cdot p^2 \cdot (1-p)^{k-2}.$$

using R to plot, $p = 1/9$, $r = 2$.

```
k <- 2:50
pk <- (k-1)·p^2·(1-p)^(k-2)
plot(k, pk, type="h")
```

The Hypergeometric Distribution

Suppose that an urn contains n balls, of which r are black and $n-r$ white. Let X denote the number of black balls drawn when taking m balls without replacement.

$$P(X=k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}},$$

X is called a hypergeometric r.v. with parameter r, n and m.

Expl. As in Expl. G of section 1.4.2, a player in the California lottery chooses 6 numbers from 53 and the lottery officials later choose 6 number at random. Let X equal the number of matches. Then

$$P(X=k) = \frac{\binom{6}{k}\binom{47}{6-k}}{\binom{53}{6}}.$$

e.g. $k=2$.

$$P(X=2) = \frac{\binom{6}{2}\binom{47}{4}}{\binom{53}{6}} = \frac{\frac{6\times 5}{2} \cdot \frac{47\times 46\times 45\times 44}{4!}}{\frac{53\cdot 52\cdot 51\cdot 50\cdot 49\cdot 48}{6!}} = 0.117.$$

k	0	1	2	3	4	5	6
$p(k)$	0.468	0.401	0.117	0.014	0.000706	0.0000122	4.36×10^{-8}

The Poisson Distribution

$X \sim P(\lambda)$, if $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, 3, \cdots$

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

In R,

```
λ <- 5
k <- 0:20
pk <- dpois(k, λ)
plot(k, pk, type="h").
```

Connection with binomial. $X \sim B(n,p)$, $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$.

using $\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$, show $p(k) \to \frac{\lambda^k}{k!} e^{-\lambda}$. $\lambda = np$. $p \to 0$, but $np \to \lambda$

i.e., $X \sim B(n,p) \to P(\lambda)$. $p \to 0$, $n \to \infty$, $np \to \lambda$.

Expl. Two dice are rolled 100 times, and the number of double sixes, X, is counted. Then $X \sim B(100, 1/36)$. Since $n=100$ large, $p = 1/36 = 0.0278$ small, $X \sim P(\lambda)$, with $\lambda = n \cdot p = 2.78$. here are comparisons:

k		0	1	2	3
$P(X=k) = \binom{100}{k}\left(\frac{1}{36}\right)^k \left(\frac{35}{36}\right)^{100-k}$	dbinom(k, 100, 0.0278)	0.0596	0.1705	0.2414	0.2255
$P(X=k) = \frac{2.78^k}{k!} e^{-2.78}$	dpois(k, 2.78)	0.0620	0.1725	0.2397	0.2221

Expl. Suppose that an office receives telephone calls as a Poisson process with $\lambda = 0.5$ per min. The number of calls in a 5-min. interval follows a Poisson dist. with parameter $\omega = 5 \cdot \lambda = 5 \times 0.5 = 2.5$. Thus, the prob. of no calls in a 5-min. interval is $P(0) = \frac{2.5^0}{0!} e^{-2.5} = 0.082$. The prob. of exactly one call is $P(1) = \frac{2.5^1}{1!} e^{-2.5} = 0.205$.

Poisson process:

Let us suppose that events are occurring at certain (random) points of time, and let us assume that for some positive λ, the following assumptions hold true.

1°. The probability that exactly 1 event occurs in a given interval of length h is $\lambda h + o(h)$.

2°. The prob. that 2 or more events occur in an interval of length h is equal to $o(h)$.

3°. For any integers n, $j_1, j_2, \cdots, j_n$, and any set of n no overlapping intervals, if we define E_i to be the event that exactly j_i of the events under consideration occur in the i-th of these intervals, then events $E_1, E_2, \cdots, E_n$ are independent.

Under these assumptions, the number of events occuring in any interval of length t is a Poisson r.v. with parameter λt.

e.g. earthquake; people entering bank, gasstation, mall; death; accident;

Expl. Suppose that earthquakes occur in the west portion of U.S. in accordance with assumptions with $\lambda = 2$ and with 1 week as the unit of time.

(a). Find the prob. that at least 3 earthquakes occur during the next 2 weeks.

(b). Find the prob. distribution of the time, starting from now, until the next earthquake.

Ans: (a) Let X be the number of earthquakes during the next 2 weeks.

Then $X \sim$ Poisson with $\lambda = 2 * 2$ $(\lambda * t)$.

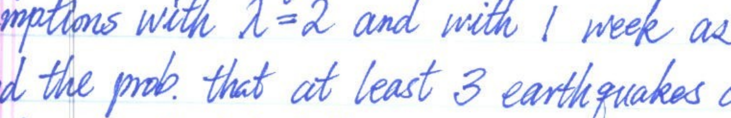
Thus $P(X \geq 3) = 1 - P(X=0,1,2)$ "complement event"

$$= 1 - P(X=0) - P(X=1) - P(X=2)$$

$$= 1 - \frac{4^0}{0!}e^{-4} - \frac{4^1}{1!}e^{-4} - \frac{4^2}{2!}e^{-4} = 0.7618$$

or $P(X \geq 3) = 1 - P(X \leq 2) = 1 - ppois(2, 4)$ in R

(b)



T = time to next earthquake

$P(T > t) = P($no earthquakes in $[0, t])$

$= P(X=0)$, where X = number of earthquakes in $[0, t]$.

then $X \sim P(\lambda)$, $\lambda = 2 * t = 2t$.

thus $P(T > t) = P(X=0) = \frac{(2t)^0}{0!}e^{-2t} = e^{-2t}$,

or $P(T \leq t) = 1 - P(T > t) = 1 - e^{-2t}$. exponential dist. r.v.

Problem 15. Two teams, A and B, play a series of games. If team A has prob. 0.4 of winning each game, is it to its advantage to play the best three out of five games or the best four out of seven? Assume the outcomes of successive games are independent.

Ans: Let X be number of games team A played to win the series.

consider 3 out of 5 games. then X follows negative binomial dist.

$P(X=k) = \binom{k-1}{3-1} 0.4^3 \cdot (1-0.4)^{k-3}$ with $r=3$, $p=0.4$.

e.g.1. $P(X=3) = 0.4^3$, $P(X=4) = \binom{3}{2}(0.4)^3(0.6)^1 = 3\cdot(0.4)^3\cdot(0.6)$.

$P(X=5) = \binom{4}{2} 0.4^3 (0.6)^2 = 6\cdot(0.4)^3\cdot(0.6)^2$.

Then, team A wins with prob $= P(X=3) + P(X=4) + P(X=5)$

$= 0.3174$

consider 4 out of 7 games. X follows negative binomial dist with $r=4$, $p=0.4$.

$P(X=k) = \binom{k-1}{4-1} 0.4^4 (1-0.4)^{k-4}$. $P(X=4) = 0.4^4 = 0.0256$

$P(X=5) = \binom{4}{3} 0.4^4\, 0.6^1 = 4\cdot 0.4^4 \cdot 0.6 = 0.06144$

$P(X=6) = \binom{5}{3} 0.4^4\, 0.6^2$ ($\to$ dnbinom(2, 4, 0.4))

$= 0.09216$

$P(X=7) =$ dnbinom(3, 4, 0.4) $= 0.110592$.

Thus, team A wins with prob: $0.0256 + 0.06144 + 0.09216 + 0.110592$

$= 0.2898$.

17. Suppose that in a sequence of independent Bernoulli trials, each with prob. of p, the number of failures up to the first success is counted. What is the frequency function for this r.v.?

Ans: Let X be the number of failures up to the first success.

then, possible values for X are $0, 1, 2, \cdots$

$P(X=0) = p$, $P(X=1) = P(F_1 S_2) = (1-p)\cdot p$.

$P(X=2) = P(F_1 F_2 S_3) = (1-p)^2 \cdot p$, $\cdots$

$P(X=k) = (1-p)^k \cdot p$, $k = 0, 1, 2, \cdots$