Special Discrete Distributions

Day 5 (1/21/20)

3.1 Binomial

- A **Bernoulli experiment** is any random experiment which results in one of two outcomes (0 or 1, success or failure, heads or tails, male or female, event or no-event).
- **Bernoulli trials** is a series of independent Bernoulli experiments
- A **Bernoulli random variable** with parameter p has pmf

x	0	1
$p(x)$	$1-p$	p

, or

$$p(x) = p^x(1-p)^{1-x},\ x = 0, 1 \tag{3.1.1}$$

- Mean, variance, and standard deviation of Bernoulli(p)
  1. $\mu = E(X) = (0)(1-p) + (1)(p) = p$
  2. $\sigma^2 = E(X-p)^2 = (0-p)^2(1-p) + (1-p)^2 p = p^2 - p^3 + p - 2p^2 + p^3 = p(1-p)$
  3. $\sigma = \sqrt{p(1-p)}$
- A **binomial random variable** is the number of successes in n Bernoulli trials.

  Example 3.1.1 Suppose we roll a die $n = 10$ times, and let Y be the number of sixes. What is, for example, $P(Y = 2)$?

  Solution: The event $Y = 2$ is the union of the *ordered* 10-tuples

$$SSFFFFFFFF \cup SFSFFFFFFF \cup \cdots \cup FFFFFFFFSS$$

$$\begin{aligned} P(Y=2) &= \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 + \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 + \cdots + \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 \\ &= \binom{10}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 \end{aligned}$$

- A random variable Y has a **binomial distribution** with parameters n and p if it has pmf

$$P(Y = j) = p(j) = \binom{n}{j} p^j (1-p)^{n-j},\ j = 0, 1, 2, \ldots, n$$

  Recall the binomial expansion

$$(a+b)^n = a^n + na^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + b^n = \sum_{j=0}^{n}\binom{n}{j}a^j b^{n-j}$$

  Then

$$\sum_{j=0}^{n} P(Y=j) = \sum_{j=0}^{n}\binom{n}{j}p^j(1-p)^{n-j} = [p + (1-p)]^n = 1$$

- To find the mean and variance of binomial(n, p),
  $M(t) = E\left(e^{tY}\right) = \sum_{j=0}^{n} e^{tj}\binom{n}{j}p^j(1-p)^{n-j} = \sum_{j=0}^{n}\binom{n}{j}\left(pe^t\right)^j(1-p)^{n-j} = \left[pe^t + (1-p)\right]^n.$
  $M'(t) = n\left[pe^t + (1-p)\right]^{n-1} pe^t$ implies
  $$E(X) = M'(0) = np$$
  $M''(t) = n\left[pe^t + (1-p)\right]^{n-1} pe^t + n(n-1)\left[pe^t + (1-p)\right]^{n-2}\left(pe^t\right)^2$ implies
  $$\mathrm{Var}(X) = M''(0) - [M'(0)]^2 = np + n(n-1)p^2 - (np)^2 = np(1-p)$$
- Binomial in R
  dbinom, pbinom, qbinom, rbinom

3.1.3 Hypergeometric Distribution

- Suppose we have a box of N items, of which D are defective and $N-D$ are nondefective. Let X be the number of defective items in a sample of size n drawn without replacement. Then X is a **hypergeometric random variable** with parameters N, D, and n.
- The pmf of X is
  $$P(X=j) = p(j) = \frac{\binom{D}{j}\binom{N-D}{n-j}}{\binom{N}{n}},\ j = 0, 1, 2, \ldots, n$$
  Example: Draw a 5-card poker hand. What is the probability exactly 2 cards are spades?
  Solution: Let X be the number of spades. Then X is hypergeometric$(52, 13, 5)$.
  $$P(X=2) = \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}}$$
- Mean and variance
  $$E(X) = n\frac{D}{N}$$
  $$\mathrm{Var}(X) = n\frac{D}{N}\frac{N-D}{N}\frac{N-n}{N-1}$$
- Hypergeometric in R (dhyper, phyper, qhyper, rhyper)

3.2 Poisson

- A random variable X has a **Poisson distribution** with parameter λ id X has pmf
  $$p(j) = \frac{e^{-\lambda}\lambda^j}{j!},\ j = 0, 1, 2, \ldots$$
  Since $1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots = e^z$, then
  $$\sum_{j=0}^{\infty} p(j) = \sum_{j=0}^{\infty}\frac{e^{-\lambda}\lambda^j}{j!} = e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = e^{-\lambda}e^{\lambda} = 1.0$$

- Mean and variance of Poisson(λ)

  $M(t) = E(e^{tX}) = \sum_{j=0}^{\infty} e^{tj} \frac{e^{-\lambda}\lambda^j}{j!} = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda e^t)^j}{j!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$

  $M'(t) = e^{\lambda(e^t-1)}\lambda e^t$ implies

  $$E(X) = M'(0) = \lambda$$

  $M''(t) = e^{\lambda(e^t-1)}\lambda e^t + e^{\lambda(e^t-1)}(\lambda e^t)^2$

  $$\text{Var}(X) = M''(0) - [M'(0)]^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda$$

- Poisson in R

  dpois, ppois, qpois, rpois

- Relationship between Poisson and Binomial

  Example: Let Y be the number of accidents in a busy intersection during the next 100 days (where the average number given past data is $\mu = 2.5$ accidents every 100 days).

  1. Case 1: Binomial with $n = 100$ days
     Let $n = 100$ days. Then $p = .025$ is the probability of one accident each day, so that $\mu = np = (100)(.025) = 2.5$. The binomial probabilities are

j	0	1	2	3	4	5	6	7	...
$p(j)$	.0795	.2039	.2588	.2168	.1348	.0664	.0269	.0093	

     (1)

  2. Case 2: Binomial with $n = 2400$ hours
     If we change the interval-lengths from days to hours, then $n = 2400$ hours and $p = 0.00104$ is the probability of one accident each hour, so that $\mu = np = (2400)(.00104) = 2.5$. The new binomial probabilities are

j	0	1	2	3	4	5	6	7	...
$p(j)$	.0820	.2052	.2566	.2139	.1337	.0668	.0278	.0099	

     (2)

  3. Case 3: Poisson
     Cases 1 and 2 illustrate an arbitrariness in our choice of unit interval (e.g. day or hour or minute, etc). As unit intervals get shorter, then n gets larger, and p shrinks because we should maintain $\mu = np = 2.5$. Equations (1) and (2) suggest the probability values may converge if we set $p = \frac{\mu}{n}$ and let $n \to \infty$.
     In general, for $\mu > 0$ and any nonnegative integer j, it can be shown that

     $$\lim_{n\to\infty} \binom{n}{j} \left(\frac{\mu}{n}\right)^j \left(1 - \frac{\mu}{n}\right)^{n-j} = \frac{e^{-\mu}\mu^j}{j!}$$

     In other words, Bin(n,p) probabilities for large n and small p are approximated by Poisson(λ) where $\lambda = np$. Here are Poisson probabilities $p(j) = \frac{e^{-2.5}2.5^j}{j!}$, $j = 0, 1, 2, 3, \ldots$

j	0	1	2	3	4	5	6	7	...
$p(j)$	.0821	.2052	.2565	.2138	.1336	.0668	.0278	.0099	

     (3)