Notes on Special Discrete Distributions

(1) Bernoulli(p) X is Bernoulli(p) if it has probability mass function given by

$$p(x) = P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Thus, $X \in \{0, 1\}$. X can be thought of as an *indicator random variable*, which indicates if some event happens ($X = 1$) or not ($X = 0$).

It is easy to compute $E(X) = 0 \cdot (1-p) + 1 \cdot p = p$. Also, $E(X^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$, so $var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$. We also get

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

(2) Binomial(n, p) X is Binomial(n, p) if its mass function is given by

$$p(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x \in \{0, 1, 2, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$$

Thus, $X \in \{0, 1, 2, \ldots, n\}$. X can be thought of as the number of *successes* in n independent trials, where the probability any one trial is a success is p.

We can compute the expectation

$$E(X) = \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i (1-p)^{n-i},$$

which evaluates to $E(X) = np$, after some manipulation. We can also compute

$$var(X) = E(X^2) - [E(X)]^2 = \sum_{i=0}^{n} i^2 \cdot \binom{n}{i} p^i (1-p)^{n-i} - [np]^2 = np(1-p)$$

(It is not obvious how to do the tricks to sum the series, but we will see another trick later for how to get $E(X)$ and $var(X)$ for a Binomial(n, p), based on observing that X can be written as a sum of n independent and identically distributed Bernoulli(p) random variables.)

One can use the mass function $p(x)$ to compute the cdf, $F(x)$, but there is no closed form expression, just an ugly summation.

(3) Poisson(λ) X is Poisson(λ) if its mass function is given by

$$p(x) = P(X = x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & \text{if } x \in \{0, 1, 2, 3, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

Thus, $X \in \{0, 1, 2, \ldots\}$. We can compute the expectation

$$E(X) = \sum_{i=0}^{\infty} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} = \lambda$$

and $var(X) = E(X^2) - [E(X)]^2$ works out to give $var(X) = \lambda$. The cdf, $F(x)$, can be obtained from $p(x)$ in the usual way, as a summation, but has no nice closed formula.

A Poisson(λ) random variable X is often used to model the number of misprints on a page of text, or the number of customers that arrive at a store in a given interval of time, or the number of "events" that occur in a given interval of time, etc.

The Poisson(λ) distribution is an approximation to the Binomial(n, p) distribution for the case that n is large, p is small, and $\lambda = np$. In other words, if Y is Binomial(n, p), and n is large and p is small, and X is Poisson(λ) with $\lambda = np$, then

$$P(Y = i) = \binom{n}{i} p^i (1-p)^{n-i} \approx P(X = i) = e^{-np} \frac{(np)^i}{i!},$$

Examples from the text: 7a (page 162), 7b (page 162), 7c (page 162), self-test problem 14, page 203.

Example: On average, there is 1 typo per 3 pages of a book. Find the probability that there is at least one typo on page 37.

Let X denote the number of typos on page 37. Since there is an average of 1 typo per 3 pages, we know that $E(X) = 1/3$. We can model X as a Poisson(1/3) random variable. We want to compute

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/3}$$

(Alternatively, $P(X \geq 1) = \sum_{i=1}^{\infty} e^{-1/3} \frac{(1/3)^i}{i!}$.)

Now suppose we are also asked to find the probability that there are at least 3 typos on page 37, given that there is at least one typo on page 37 that you have already seen. We want to compute

$$P(X \geq 3 \mid X \geq 1) = \frac{P(X \geq 3, X \geq 1)}{P(X \geq 1)} = \frac{P(X \geq 3)}{P(X \geq 1)} = \frac{1 - e^{-1/3}(1 + (1/3) + (1/3)^2/2)}{1 - e^{-1/3}}$$

Example: Suppose that the probability that a radio lasts more than 15 years is $p = e^{-3}$. (a). What is the (exact) probability that, of 1000 such radios that you purchase, at least four of them last more than 15 years?

Let X be the number (among the 1000) that last more than 15 years. Then, X is Binomial(1000,p), where $p = e^{-3}$. We want

$$P(X \geq 4) = \sum_{i=4}^{1000} \binom{1000}{i} p^i (1-p)^{1000-i}$$

(b). Give an alternative expression that approximates *your answer to part (b).*

Since 1000 is "large" and $p = e^{-3}$ is "small", we know that X has a distribution that is approximated by that of Y, a Poisson($1000p$) random variable. Thus,

$$P(X \geq 4) \approx P(Y \geq 4) = \sum_{i=4}^{\infty} e^{-1000p} \frac{(1000p)^i}{i!}$$

(Note: you could also stop the summation at $i = 1000$; it does not really matter for the approximation.)

Example: People enter a store, on average one every two minutes. (a). What is the probability that no people enter between 12:00 and 12:05?

Let X be the number of people that enter between 12:00 and 12:05. We model X as a Poisson(λ) random variable. What should λ be? It should be such that $\lambda = E(X)$, the average number of people that arrive in the 5-minute interval. Well, if 1 person arrives every 2 minutes, on average (so 1/2 a person per minute), in 5 minutes and average of 2.5 people will arrive. Thus, $\lambda = 2.5$.

We want to compute $P(X = 0) = e^{-2.5} \frac{(2.5)^0}{0!} = e^{-2.5}$.

(b). Find the probability that at least 4 people enter during [12:00,12:05].

We want

$$P(X \geq 4) = 1 - e^{-2.5} - \frac{(2.5)^1}{1!} e^{-2.5} - \frac{(2.5)^2}{2!} e^{-2.5} - \frac{(2.5)^3}{3!} e^{-2.5}$$

---

(4) Geometric(p) X is Geometric(p) if its mass function is given by

$$p(x) = P(X = x) = \begin{cases} (1-p)^{x-1} \cdot p & \text{if } x \in \{1, 2, 3, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

Thus, $X \in \{1, 2, \ldots\}$. We can compute the expectation

$$E(X) = \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} \cdot p = \frac{1}{p}$$

and $var(X) = E(X^2) - [E(X)]^2$ works out to give $var(X) = (1-p)/p^2$.

The cdf, $F(x)$, can be obtained from $p(x)$ in the usual way, as a summation, and actually does have a nice closed formula. For $i = 1, 2, 3, \ldots$, we get

$$P(X \leq i) = 1 - P(X \geq i+1) = 1 - \sum_{j=i+1}^{\infty} (1-p)^{j-1} \cdot p = 1 - p(1-p)^i \sum_{j=0}^{\infty} (1-p)^j = 1 - (1-p)^i.$$

Thus, the cdf is

$$F(x) = P(X \le x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 - (1-p)^i & \text{if } i \le x < i+1, \text{ for } i = 1, 2, 3, \ldots \end{cases}$$

A Geometric(p) random variable X can be thought of as the number of trials until first success when independent trials are repeated over and over until the first success (up to and including the success; thus, $X \ge 1$).

Example: Suppose parts are of two varieties: good (with probability 90/92) and slightly defective (with probability 2/92). Parts are produced one after the other. What is the probability that at least 5 parts must be produced until there is a slightly defective part produced?

Let X be the number of parts produced up to (and including) the first slightly defective part. Then, X is Geometric(2/92).

We want to compute

$$P(X \ge 5) = \sum_{i=5}^{\infty} (90/92)^{i-1}(2/92) = \frac{2}{92}\left(\frac{90}{92}\right)^4 \cdot \frac{1}{1 - \frac{90}{92}} = \left(\frac{90}{92}\right)^4$$

(Alternatively, we could have written $P(X \ge 5) = 1 - P(1 \le X \le 4)$, and written out a short summation of terms; it gives the same answer, of course.)

---

(5) Negative Binomial (r, p) X is NegBin(r, p) if its mass function is given by

$$p(x) = P(X = x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{if } x \in \{r, r+1, r+2, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

Here, r is a positive integer and $0 < p < 1$. Thus, $X \in \{r, r+1, r+2, \ldots\}$. (A NegBin random variable is also known as a "Pascal" random variable.)

We can compute the expectation

$$E(X) = \sum_{i=1}^{\infty} i \cdot \binom{i-1}{r-1} p^r (1-p)^{i-r} = \frac{r}{p}$$

and $var(X) = E(X^2) - [E(X)]^2$ works out to give $var(X) = r(1-p)/p^2$. The cdf, $F(x)$, can be obtained from $p(x)$ in the usual way, as a summation, but has no nice closed formula.

A NegBin(r, p) random variable X can be thought of as the number of trials until the rth success when independent trials are repeated over and over until the rth success (up to and including the rth success; thus, $X \ge r$).

Example: [problem 64, page 195, Ross] The suicide rate in a certain state is 1 suicide per 100,000 inhabitants per month. (a). Find the probability that in a city of 400,000 inhabitants within this state, there will be 8 or more suicides in the month of June this year.

Let X be the number of suicides in this city in the month of June this year. Then, X is counting the number of "successes" in 400,000 independent trials, where "success" here means that a person commits suicide during the month of June, which happens with probability 1/100,000. We are making the assumption of independence: one person's act of suicide does not influence another person to commit (or not) suicide. Thus, X is Binomial(n, p), where $n = 400,000$, $p = 1/100,000$.

We want to compute

$$\alpha = P(X \ge 8) = 1 - P(X \le 7) = 1 - \left(\binom{n}{0}p^0(1-p)^{n-0} + \binom{n}{1}p^1(1-p)^{n-1} + \cdots + \binom{n}{7}p^7(1-p)^{n-7}\right)$$

Now, we can also make a very good estimate of this probability, since $n = 400,000$ is large and $p = 1/100,000$ is small: X is approximately Poisson(λ), where $\lambda = 400,000(1/100,000) = 4$. (Note that X is **not** a Poisson random variable; we are just using it as an approximation.) Thus,

$$\alpha = P(X \ge 8) = 1 - P(X \le 7) = 1 - \left(1 + \frac{4}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!} + \cdots + \frac{4^7}{7!}\right)e^{-4}$$

(b). What is the probability that there will be at least two months during the year in which the city has 8 or more suicides?

Let Y be the number of months this year in which the city has 8 or more suicides (a "bad month"). Then $Y \in \{0, 1, 2, \ldots, 12\}$, and, assuming independence among the months, Y is Binomial$(12, \alpha)$, where α is the number from part (a). We want to compute

$$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - \binom{12}{0}\alpha^0(1-\alpha)^{12} - \binom{12}{1}\alpha^1(1-\alpha)^{11}.$$

(c). Counting the present month as month number 1, what is the probability that the first month to have 8 or more suicides (in the city) will be month number i, $i \geq 1$?

Let Z be the number of the first month that has 8 or more suicides. Then Z is Geometric(α). We want to compute $P(Z = i) = (1-\alpha)^{i-1}\alpha$, for $i = 1, 2, 3, \ldots$.

(d). Today is January 1. What is the probability that the fourth month to have 8 or more suicides in the city is December?

Let U be the month number of the fourth month to have 8 or more suicides in the city. Then U is NegBin$(4, \alpha)$, since it counts how many "trials" (months) until the 4th "success" (a bad month, with 8 or more suicides). We want to compute

$$P(U = 12) = \binom{12-1}{4-1}\alpha^4(1-\alpha)^{12-4}$$
