Weekly Notes for EE2012 2014/15 – Week 6

T. J. Lim

February 18, 2015

Book sections covered this week: 3.5, 4.1.

1 Important Discrete Random Variables

Discrete r.v.s often represent a count of the number of occurrences of some random phenomenon, and certain ones appear again and again in many contexts due to the prevalence of certain types of counting problems in many applications.

1.1 Bernoulli Distribution

The Bernoulli r.v. I can take only two values 0 or 1, with the pmf $p_I(0) = 1 - p$, $p_I(1) = p$, where $0 < p < 1$. Its mean and variance are

$$E[I] = p \tag{1}$$
$$\sigma_I^2 = p(1-p) = pq \tag{2}$$

where $q = 1 - p$.

It is usually employed to indicate whether an event of interest, e.g. failure of component, arrival of a customer, occurred.

1.2 Binomial Distribution

The binomial r.v. X is the number of successes in a fixed number n of independent and identical[1] Bernoulli trials. As we have discussed several times before, the pmf of X is $p_X(k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \ldots, n$ where p is the probability of succeeding in any one Bernoulli trial and $q = 1 - p$.

The mean and variance can be computed by brute force with some difficulty, as shown in Examples 3.27 and 3.28 of the textbook. We will encounter a much simpler method for computing these values later and so only present the final results here:

$$\mu_X = np, \quad \sigma_X^2 = npq. \tag{3}$$

[1] In other words, each trial has the same probability of success, p.

1.3 Geometric Distribution

The geometric r.v. M is the number of Bernoulli trials required up to and including the first success[2], and has the pmf

$$p_M(k) = (1-p)^{k-1}p, \quad k = 1, 2, \dots \tag{4}$$

The mean and variance can be found from first principles as shown in Examples 3.15 and 3.22 in the textbook. As with the binomial case, there are easier ways with which to calculate the mean and variance which we will discuss later in the course. The final results are

$$\mu_M = \frac{1}{p}, \quad \sigma_M^2 = \frac{q}{p^2}. \tag{5}$$

The geometric random variable is relatively easy to work with, due to the geometric nature of its pmf. For instance,

$$P[M > k] = \sum_{j=k+1}^{\infty} q^{j-1}p = q^k, \tag{6}$$

meaning that the probability of waiting longer than k Bernoulli trials for your first success decreases exponentially with k. The rate of decay depends on $q = 1 - p$ – the smaller the q, the faster the decay i.e. the less likely you will have to wait a long time.

The geometric r.v. is also the only one with $S_X = \{1, 2, \dots\}$ that has the *memory-less property*:

$$P[M \geq k + j | M > j] = P[M \geq k]. \tag{7}$$

In words, the probability of waiting a further k time intervals for your first success given you've already waited j time intervals, is the same for all j. So unfortunately, the chances of winning the lottery do not increase with the number of attempts. Whether you have already played 100 times or are buying your first ticket, you will have the same probability of winning. This is quite logical given the independence of the sequence of Bernoulli trials.

1.4 Poisson Distribution

The Poisson random variable N with parameter $\alpha > 0$ has the following pmf:

$$p_N(k) = \frac{\alpha^k}{k!}e^{-\alpha}, \quad k = 0, 1, 2, \dots \tag{8}$$

[2] It can also be defined as the number of failures before the first success but for consistency, we use the definition that includes the one success throughout this course.

Before discussing its importance, we first derive its mean and variance:

$$\begin{aligned}
\mu_N &= \sum_{k=0}^{\infty} k \frac{\alpha^k}{k!} e^{-\alpha} && (9)\\
&= e^{-\alpha} \sum_{k=1}^{\infty} \frac{\alpha^k}{(k-1)!} && (10)\\
&= e^{-\alpha} \sum_{l=0}^{\infty} \frac{\alpha^{l+1}}{l!} && (11)\\
&= \alpha e^{-\alpha} \sum_{l=0}^{\infty} \frac{\alpha^l}{l!}. && (12)
\end{aligned}$$

But the power series expansion of e^x is

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad x \in \mathbb{C}. \tag{13}$$

Therefore, we have

$$\mu_N = \alpha e^{-\alpha} e^{\alpha} = \alpha. \tag{14}$$

To find the variance, we can first find $E[N^2]$.

$$\begin{aligned}
E[N^2] &= \sum_{k=0}^{\infty} k^2 \frac{\alpha^k}{k!} e^{-\alpha} && (15)\\
&= e^{-\alpha} \sum_{k=1}^{\infty} k \frac{\alpha^k}{(k-1)!} && (16)\\
&= e^{-\alpha} \sum_{l=0}^{\infty} (l+1) \frac{\alpha^{l+1}}{l!} && (17)\\
&= \alpha e^{-\alpha} \left(\sum_{l=0}^{\infty} l \frac{\alpha^l}{l!} + \sum_{l=0}^{\infty} \frac{\alpha^l}{l!} \right). && (18)
\end{aligned}$$

But

$$e^{-\alpha} \sum_{l=0}^{\infty} l \frac{\alpha^l}{l!} = E[N] = \alpha, \quad \text{and} \quad \sum_{l=0}^{\infty} \frac{\alpha^l}{l!} = e^{\alpha}, \tag{19}$$

and hence

$$E[N^2] = \alpha^2 + \alpha.$$

Finally, $\sigma_N^2 = E[N^2] - \mu_N^2 = \alpha$. Therefore the mean and variance of a Poisson random variable are both equal to α.

1.4.1 Importance and Derivation of Poisson Distribution

Consider a finite time interval of T seconds (or any other unit of time). Within this interval, there could be a number of independent occurrences of some random event of interest, e.g. customers coming into a store, data packets arriving at a transmitter, radioactive particles emitted. The Poisson r.v. N models the number of such occurrences over a fixed period of time T *provided* that they occur independently of each other, and only one at a time. We now explain why.

Let the time interval $[0, T]$ be divided into n equal sub-intervals of duration T/n. If T/n is large (i.e. n is small), then in any sub-interval k there is a non-zero probability of seeing 0, 1, 2, or more events of interest. But as n grows without bound, i.e. we let $n \to \infty$, then the probability of seeing more than one event occurrence in sub-interval k vanishes. So the number of events in each sub-interval is either 0 or 1, a Bernoulli random variable with mean value p, as yet unknown. The total number of such events within $[0, T]$ is now seen to be binomial, with parameters n and p.

Now the average number of these events is given as α, and by the binomial argument above, we then have $\alpha = np$, or $p = \alpha/n$. The pmf of N, the total number of events in $[0, T]$, can thus be written as

$$p_N(k) = \binom{n}{k} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k}, \quad k = 0, 1, 2, \ldots \tag{20}$$

where the range of N has become infinite because we have taken the limit $n \to \infty$.

Now we make use of the fact that

$$\lim_{n\to\infty} \left(1 - \frac{\alpha}{n}\right)^n = e^{-\alpha}, \tag{21}$$

which is sometimes used as the definition of the exponential function and can be verified through a comparison of the terms within the power series expansions on both sides, to obtain

$$p_N(k) = \frac{n!}{(n-k)!k!} \left(\frac{\alpha/n}{1 - \alpha/n}\right)^k e^{-\alpha}. \tag{22}$$

Since n is large compared to any finite α, we can set $1 - \alpha/n$ to 1, so that

$$p_N(k) = \frac{\prod_{j=0}^{k-1}(n-j)}{k!} \frac{\alpha^k}{n^k} e^{-\alpha}. \tag{23}$$

Finally, for any finite value of k, $\lim_{n\to\infty} \prod_{j=0}^{k-1}(n-j) = n^k$, and hence we have

$$p_N(k) = \frac{\alpha^k}{k!} e^{-\alpha}, \quad k = 0, 1, 2, \ldots \tag{24}$$

Poisson approximation of binomial A small change to the derivation above (see pg 137 of the book) reveals that a binomial random variable with large n and small p, with finite mean value np, can be approximated by a Poisson r.v. with parameter $\alpha = np$. In other words,

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{(np)^k}{k!} e^{-np} \tag{25}$$

when $n \to \infty$ and $p \to 0$, and $np < \infty$. Such an approximation is practically useful because $\binom{n}{k}$ for very large n cannot be computed.

Example 1: Assume that the probability of a living person having been hit by a vehicle up to this point in his/her life is 1 in 1 million. Among the 5 million people currently living in Singapore, what is the probability that fewer than 10 have ever been hit by a car?

The number of living people in Singapore who have been hit by a car, denoted by N, is binomial with $n = 5 \times 10^6$ and $p = 1 \times 10^{-6}$. We need to find $\sum_{k=0}^{9} p_N(k)$, but $\binom{n}{k}$ with $n = 5 \times 10^6$ is impossible to compute. Instead we use the Poisson appoximation:

$$\sum_{k=0}^{9} p_N(k) \approx e^{-5} \sum_{k=0}^{9} \frac{5^k}{k!} = 0.9682. \tag{26}$$

So with 97 percent probability, fewer than 10 people in Singapore have ever been hit by a car and are still surviving today, assuming the values quoted above are believable. ■

Example 2: Customers arrive at a store singly and independently[3], at an average rate of $\lambda = 5$ customers per hour. Each customer spends an average of \$50 in the store. Find the average store revenue in a two-hour period.

Let the number of customers in two hours be N. Then N is Poisson with parameter $\alpha = 2\lambda = 10$. The average store revenue in two hours is then $50E[N]$ dollars[4]. Since $E[N] = \alpha = 10$, the average revenue in two hours is \$500. ■

In the latter example, notice that the *rate* λ of customer arrivals is given. This is a common and sensible practice, which allows us to find the distribution of the number of event occurrences over an arbitrary period of time $[0, t]$, $N(t)$:

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \ldots \tag{27}$$

1.5 Uniform Distribution

The uniform discrete r.v. has equal probability of taking each value in its sample space, i.e. if $|S_X| = L$, then $p_X(x) = \frac{1}{L}$, for all $x \in S_X$. Of particular interest is the case of S_X being a set of consecutive integers $\{j+1, \ldots, j+L\}$, for some $j \in \mathbb{Z}$. In this case,

$$\begin{aligned} E[X] &= \sum_{k=1}^{L} (j+k) \frac{1}{L} && (28) \\ &= j + \frac{1}{L} \sum_{k=1}^{L} k && (29) \\ &= j + \frac{L+1}{2}, && (30) \end{aligned}$$

and

$$\sigma_X^2 = \sum_{k=1}^{L} \left(k - \frac{L+1}{2}\right)^2 \frac{1}{L}. \tag{31}$$

[3]The Poisson model does not work if customers arrive in groups, or influence each other's arrival times.

[4]Try to derive this using conditional expectation.

Using the identity (check "Summation" in Wikipedia)

$$\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6},$$

after some manipulations, we have

$$\sigma_X^2 = \frac{L^2-1}{12}. \tag{32}$$

2 Cumulative Distribution Function (CDF)

We have seen how the probability mass function $p_X(x)$ characterizes a discrete random variable fully, in the sense that $p_X(x)$ allows us to compute the probability of any event involving X. However, for random variables with ranges that include a continuum of values, such as when X is drawn from the set of real numbers between 0 and 1, $P[X=x]$ for some or all values of $x \in S_X$ may be zero. Therefore, we need another way of describing the randomness in X that applies to such random variables.

2.1 Definition

The cumulative distribution function (CDF) of a r.v. X is defined as

$$F_X(x) = P[X \le x] = P[\{\zeta : X(\zeta) \le x\}] \tag{33}$$

where the last expression reminds us that $P[X \le x]$ is really a shorthand for an event within the underlying probability space. In words, it is the probability of obtaining an outcome ζ such that $X(\zeta)$ is not larger than x.

The CDF is defined for all types of random variables (not only the discrete sort), as shown in the examples below.

Example 3: Consider a binomial r.v. X with $n=3$ and $p=0.5$. The PMF of X is

$$p_X(0) = p_X(3) = \frac{1}{8}, \quad p_X(1) = p_X(2) = \frac{3}{8}. \tag{34}$$

The CDF of X is obtained by considering the value of $F_X(x)$ within certain critical ranges:

$$\begin{aligned}
x < 0 \quad &\Rightarrow \quad F_X(x) = 0 \\
0 \le x < 1 \quad &\Rightarrow \quad F_X(x) = P[X=0] = \frac{1}{8} \\
1 \le x < 2 \quad &\Rightarrow \quad F_X(x) = P[X=0] + P[X=1] = \frac{1}{2} \\
2 \le x < 3 \quad &\Rightarrow \quad F_X(x) = \sum_{k=0}^{2} P[X=k] = \frac{7}{8} \\
x \ge 3 \quad &\Rightarrow \quad F_X(x) = 1.
\end{aligned}$$

Therefore $F_X(x)$ is a non-decreasing staircase function. By defining the unit step function as

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \ge 0 \end{cases} \tag{35}$$

we can write compactly

$$F_X(x) = \sum_{k=0}^{3} p_X(k)u(x-k). \tag{36}$$

Example 4: Suppose a pointer is spun, and the resting position of the tip noted. Let Θ be the angle the pointer makes with the positive horizontal axis, and define $\Theta \in [0, 2\pi)$. What is the CDF of Θ?

Assuming that the pointer is not biased to rest in any particular direction, for $0 \leq \theta_1 \leq \theta_2 < 2\pi$, we have

$$P[\theta_1 < \Theta \leq \theta_2] = \frac{\theta_2 - \theta_1}{2\pi}. \tag{37}$$

To derive $F_\Theta(\theta)$, we again divide $\mathbb{R}$ into several regions, in this case $(-\infty, 0)$, $[0, 2\pi)$, $[2\pi, \infty)$ according to the range of Θ. It is easy to find the CDF in the first and third regions:

$$\begin{aligned} \theta < 0 &\Rightarrow F_\Theta(\theta) = 0 \\ \theta \geq 2\pi &\Rightarrow F_\Theta(\theta) = 1. \end{aligned}$$

For the second region, we note that $\{\Theta \leq \theta\} = \{\Theta < 0\} \cup \{0 \leq \Theta \leq \theta\}$ and let $\theta_1 = 0$ and $\theta_2 = \theta$ in (37), to obtain

$$F_\Theta(\theta) = P[\Theta < 0] + P[0 \leq \Theta \leq \theta] = \frac{\theta}{2\pi}. \tag{38}$$

Thus the CDF of Θ is piece-wise linear, with a slope of $\frac{1}{2\pi}$ between $\theta = 0$ and $\theta = 2\pi$, and flat everywhere else. We say that Θ is uniform in $[0, 2\pi)$.

Example 5: Let the waiting time for a taxi at a particular taxi stand be T hours. The event $\{T = 0\}$ occurs when no passengers are waiting in line and one or more taxis are at the stand. Let $P[T = 0] = p > 0$. Given that $T > 0$, i.e. you have to wait, then T is uniform in the range $(0, 1]$. To find the CDF of T, we can use the theorem on total probability:

$$F_T(t) = P[T \leq t | T = 0]p + P[T \leq t | T > 0](1 - p). \tag{39}$$

It should be clear that

$$P[T \leq t | T = 0] = u(t), \tag{40}$$

where $u(t)$ is the unit step. Also, we know that T is uniform in $(0, 1]$ when we have to wait, and hence

$$P[T \leq t | T > 0] = \begin{cases} 0, & t \leq 0 \\ t, & 0 < t \leq 1 \\ 1, & t > 1 \end{cases} \tag{41}$$

as in the previous example. Substituting into (39), we have

$$F_T(t) = \begin{cases} 0, & t < 0 \\ p + (1-p)t, & 0 \leq t \leq 1 \\ 1, & t > 1 \end{cases}. \tag{42}$$

Examples 3, 4 and 5 introduce three different types of random variables – discrete, continuous and mixed, respectively. Now that we know about the CDF, we can use the CDF of a r.v. to classify it into the three categories, namely:

Discrete If a r.v. has a staircase function for a CDF, it is a discrete r.v.

Continuous If a r.v. has a continuous function for a CDF, with no jump discontinuities but possibly discontinuities in its derivatives, it is a continuous r.v.

Mixed If a r.v. has a CDF that has jump discontinuities but is not a staircase function, then it is a mixed random variable.

2.2 Properties of the CDF

The CDF is a probability value, and hence we have **Property 1**

$$0 \leq F_X(x) \leq 1, \quad \forall\, x \in \mathbb{R}. \tag{43}$$

Next, because $\lim_{x\to\infty}\{X \leq x\}$ is the certain event, we get **Property 2**

$$\lim_{x\to\infty} F_X(x) = 1. \tag{44}$$

Then because $\lim_{x\to-\infty}\{X \leq x\}$ is the impossible event, we get **Property 3**

$$\lim_{x\to-\infty} F_X(x) = 0. \tag{45}$$

Consider two constants a and b such that $a < b$. Then

$$(-\infty, b] = (-\infty, a] \cup (a, b]. \tag{46}$$

By Axiom III, we then have

$$F_X(b) = F_X(a) + P[a < X \leq b]. \tag{47}$$

Since $P[a < X \leq b] \geq 0$, we have **Property 4**

$$a < b \Rightarrow F_X(a) \leq F_X(b), \tag{48}$$

or in other words, $F_X(x)$ is a non-decreasing function of x.

From (47), we also get **Property 5**

$$P[a < X \leq b] = F_X(b) - F_X(a). \tag{49}$$

Property 5 means that $F_X(x)$ can be used to calculate the probability of any event involving X, and the CDF is therefore a complete characterization of a random variable.

Property 6 states that any CDF is continuous from the right:

$$\lim_{h\to 0^+} F_X(x+h) = F_X(x). \tag{50}$$

This property can only be proven using more sophisticated tools than what we have studied, and so we have to accept that it is true. An example of a CDF that is continuous from the right, but not from the left, is the CDF of a binomial r.v. as shown in Example 3 above.

A CDF can have jump discontinuities, with the size of the jump at $x = x_i$ being equal to $P[X = x_i]$. This can be stated as **Property 7**

$$P[X = b] = F_X(b) - F_X(b^-) \tag{51}$$

where b^- is a value vanishingly close to b but smaller. Again, we can look to the discrete RV example to see that this is not so surprising.

3 Diagnostic Questions

1. Verify the memory-less property of the geometric random variable by deriving $P[M \geq k + j | M > j]$.

2. A quiet road has an average of 30 cars per hour passing through. Suppose we model the number of cars per hour as a Poisson random variable. Find the probability that in ten minutes, fewer than 3 cars pass through.

3. Let X be uniform in $\{-2, -1, 0, 1, 2\}$. Sketch the CDF of X.

4. Using Property 7 of the CDF, show that for any continuous r.v. X, $P[X = a] = 0$ for any $a \in \mathbb{R}$.