Lecture 2 The Involvement of Bernoulli Random Variables in Discrete PDFs

The Bernoulli Random Variable- The random variable *X* with sample space $S_X = \{0,1\}$ and with probability density function (pdf) $f_X(x) = \Pr[X = x] = \begin{cases} 1-p & for\ x=0 \\ p & for\ x=1 \end{cases}$ is said to be a *Ber(p)* random variable.

[See also: https://en.wikipedia.org/wiki/Bernoulli_distribution]

IMPORTANT NOTE: Often, a collection $Ber(p)$ random variables is denoted as $\{X_k\}_{k=1}^n$. Many books use this notation, and we, ourselves, will use it at times. However, in this set of notes we will denote this collection as an *ordered n-tuple* $\mathbf{X} = (X_1, X_2, \cdots, X_n)$. We choose this notation here because the notion of *ordering* is central in relation to understanding the basis for the many random variables discussed here and in Chapter 4 of the book.

At a more technical level, using the notation $\{X_k\}_{k=1}^n$ diminishes an appreciation of the importance of having a well-defined *sample space*. On the other hand, given $\mathbf{X} = (X_1, X_2, \cdots, X_n)$, one should (almost immediately) be able to identify the sample space as $S_{\mathbf{X}} = \{(x_1, x_2, \cdots, x_n) \mid x_k \in \{0,1\}\ for\ each\ k\}$. Having this description, when one is asked, for example, to identify the subset of $S_{\mathbf{X}}$ that corresponds to the event $[Y=1]$, [where $Y = \sum_{k=1}^{n} X_k$] as:

$$A_1 \stackrel{\Delta}{=} \{ (1,0,\cdots,0)\ ,\ (0,1,0,\cdots,0)\ ,\ \cdots\ ,\ (0,0,\cdots,0,1)\ \}. \tag{1}$$

Because this set includes *n* distinct elements, each of which has probability equal to $p^1(1-p)^{n-1}$, it becomes immediately apparent that:

$$\Pr[Y=1] = \Pr(A_1) = \binom{n}{1} p^1 (1-p)^{n-1}. \tag{2}$$

Consequently, it is easier to appreciate the generalization of (2) that is the *binomial pdf*:

$$\Pr[Y=y] = \binom{n}{y} p^y (1-p)^{n-y} \quad for\ \ y \in \{0,1,\cdots,n\}. \tag{3}$$

Many students who are simply given (3) have no idea of its origin, and can be easily intimidated by the inclusion of the 'choose' term, $\binom{n}{y} \stackrel{\Delta}{=} \frac{n!}{y!(n-y)!}$. In such instances, they revert to simply memorizing (3).

Remark 1. I find it interesting that the authors would place the formal development of the concepts of multiple random variables, *independence* and *conditional probability* in Chapter 5 of their book, since, as we shall see, those concepts are needed for a conceptual understanding of many of the random variables presented in Chapter 4.

We will now summarize how various popular discrete random variables are related to a collection of Bernoulli random variables.

The Binomial Random Variable- [https://en.wikipedia.org/wiki/Binomial_distribution] In statistical jargon: Suppose we conduct n trials, and that each trial can be considered to be a success or a failure. Suppose further that (A1) these trials are mutually *independent*, and that (A2) the probability of success in any given trial is p. Let Y denote the total number of successes. Then Y has a binomial pdf with parameters n and p. In the jargon of random variables: Let $\mathbf{X}=(X_1, X_2, \cdots, X_n)$ be a collection of **independent and identically distributed (*iid*)** *Ber*(p) random variables. Then $Y=\sum_{k=1}^{n} X_k$ is a *bino*(n,p) random variable. The pdf for Y is: $f_Y(y)=\binom{n}{y}p^y(1-p)^{n-y}$ on $S_Y=\{0,1,2,\cdots,n\}$ where $\binom{n}{y}\overset{\Delta}{=}\frac{n!}{y!(n-y)!}$ is stated, in words, as 'n choose y'. [https://en.wikipedia.org/wiki/Binomial_coefficient]. In lay terms, it is the number of ways that one can position y indistinguishable objects in n ordered slots. For example, there are n ways that one can position a 1 in n slots (i.e. 1st position, **or** 2nd position, **or**, … , **or** n^{th} position). In this case $\binom{n}{1}\overset{\Delta}{=}\frac{n!}{1!(n-1)!}=n$. The word **or** here is in bold to emphasize the fact that it is, indeed, a *union* set operation (denoted as $\cup$). Consider the event $[Y=1]$. In relation to the sample space $S_{\mathbf{X}}$ for $\mathbf{X}=(X_1, X_2, \cdots, X_n)$, this event is: $[Y=1]\leftrightarrow\{(1,0,\cdots,0),(0,1,\cdots,0),\cdots(0,0,\cdots,1)\}$. There are a total of n ordered *n-tuples* in this set. The first *n-tuple* has probability $\Pr\{(1,0,\cdots,0)\}\overset{\Delta}{=}\Pr[X_1=1\cap X_2=0\cap\cdots X_n=0]$. Because of *mutual independence* we have $\Pr\{(1,0,\cdots,0)\}=\Pr[X_1=1]\Pr[X_2=0]\cdots\Pr[X_n=0]\overset{\Delta}{=}p_1(1-p_2)\cdots(1-p_n)$. Because these random variables are *identically distributed*, this becomes $\Pr\{(1,0,\cdots,0)\}=p(1-p)\cdots(1-p)=p(1-p)^{n-1}$.

[Related *Matlab* commands: binopdf, binocdf, binoinv, binornd]

<u>NOTE:</u> The *Ber(p)* random variable is a special case of a *bino*(n=1,p) random variable.

Example You are concerned with X=*The act of noting whether or not any chosen part conforms to specifications.* Let $\Pr[X=1]=p$ denote the probability that it does not conform. You will inspect $n=10$ randomly selected parts. Let $\mathbf{X}=(X_1, X_2, \cdots, X_{10})$ denote the 10-D data collection variables associated with X. Then $Y=\sum_{k=1}^{10} X_k \sim bino(n=10, p)$.

(a) Assuming the truth model $p=0.05$, compute $\Pr[Y\le 1]$. <u>*Solution*</u>: $\Pr[Y\le 1]=$ binocdf(1,10,.05) = 0.9139.

(b) For $p=0.05$ simulate 5 measurements of $\hat{p}=Y/10$.
<u>*Solution*</u>: phat=binornd(10,.05,1,5)/10 phat = 0 0.2000 0.1000 0.1000 0

(c) Explain why your estimator in (b) will <u>never</u> result in the estimate 0.05 that is the true value of p.
<u>*Explanation*</u>: The sample space for $\hat{p}=Y/10$ is $S_{\hat{p}}\{0, 0.1, 0.2, \cdots, 0.9, 1\}$.

The Geometric Random Variable- [https://en.wikipedia.org/wiki/Geometric_distribution] In statistical jargon: Suppose we conduct repeated trials and that each trial can be considered to be a success or a failure. Suppose further that (A1) these trials are mutually *independent*, and that (A2) the probability of success in any given trial is *p*. Let *Y* denote the number of the trial that the first success occurs. Then *Y* has a geometric pdf with parameter *p*.
In the jargon of random variables: Let $\mathbf{X} = (X_1, X_2, \cdots, X_\infty)$ be an ordered ∞-*tuple* of *iid Ber*(*p*) random variables. Then $Y = \min_k [X_k = 1]$ is a *geo*(*p*) random variable. The pdf for *Y* is: $f_Y(y) = p(1-p)^{y-1}$ on $S_Y = \{1, 2, \cdots, \infty\}$. Consider the event $[Y=3] \leftrightarrow \{(0, 0, 1, *, \cdots, *) = [X_1 = 0 \cap X_2 = 0 \cap X_3 = 1 \cap X_4 = * \cdots, X_\infty = *]$ where * denotes 'anything'. Because of *mutual independence* we have:

$\Pr[Y=3] = \Pr[X_1 = 0]\Pr[X_2 = 0]\ \Pr[X_3 = 1]\ \Pr[X_4 = *] \cdots \Pr[X_\infty = *] = (1-p_1)(1-p_2)p_3 1 \times \cdots \times 1 = (1-p_1)(1-p_2)p_3$. Because they are *identically distributed*, we arrive at $\Pr[Y=3] = p(1-p)^2$
[Related *Matlab* commands: geopdf, geocdf, geoinv, geornd]

<u>NOTE:</u> The *geo*(*p*) random variable is the act of measuring the first trial at which a success occurs in a sequence of Bernoulli trials. However, in *Matlab* it is defined as the number of failures *prior* to the first success. It should be clear then that the event $[Y = y]$ is the same as Matlab's event $[Y_{Matlab} = y - 1]$

Example Suppose that on any occasion you go out, the probability that you get to know someone new is $\Pr[X = 1] = 0.1$.

(a) What is the probability that you will need to go out exactly 5 times in order to meet a new person?
<u>*Answer*</u>: $\Pr[Y = 5] = 0.9^4(0.1) = 0.0656$ = geopdf(4,0.1)

(b) What is the probability that you will need to go out no more than 5 times in order to meet a new person?
<u>*Answer*</u>: >> geocdf(4,0.1) = 0.4095

(c) What is the maximum number of times you would need to go out in order to have a 90% chance of meeting someone new?
Answer: geoinv(.9,.1) = 21

The Negative Binomial Random Variable- [https://en.wikipedia.org/wiki/Negative_binomial_distribution] In statistical jargon: Suppose we conduct repeated trials and that each trial can be considered to be a success or a failure. Suppose further that (A1) these trials are mutually *independent*, and that (A2) the probability of success in any given trial is *p*. Let *Y* denote the number of the trial that the r^{th} success occurs. Then *Y* has a negative binomial pdf with parameters *r* and *p*.

In the jargon of random variables: Let $\mathbf{X} = (X_1, X_2, \cdots, X_\infty)$ be an ordered ∞-*tuple* of *iid Ber*(*p*) random variables. Then $Y = \min_n \sum_{k=1}^{n} X_k = r$ is a *nbino*(*r*,*p*) random variable. The pdf for *Y* is: $f_Y(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$ on $S_Y = \{r, r+1, r+2, \cdots, \infty\}$.

[Related *Matlab* commands: nbinpdf, nbincdf, nbininv, nbinrnd]

NOTE: The *nbino*(*r*,*p*) random variable is the act of measuring the trial at which the rth success occurs in a sequence of Bernoulli trials. However, in *Matlab* it is defined as the number of *extra trials prior to* the rth success. It should be clear then that the event $[Y = y]$ is the same as Matlab's event $[Y_{Matlab} = y - r]$

Example In reforestation of a particular species of tree it is known that the probability that any given tree will 'take root' is $\Pr[X = 1] = 0.8$. For a given tract of land, you want to grow $r = 20$ trees.

(a) What is the probability that you will get $r = 20$ by planting only exactly $y = 25$ trees.

Solution: $f_Y(25) = \binom{25-1}{20-1}(0.8^{20})(0.2^5)$ =nchoosek(24,19)*(0.8^20)*(0.2^5) = 0.1568

We can also use: nbinpdf(5,20,.8) = 0.1568 (i.e. we need 5 extra trials)

(b) What is the probability that you will get $r = 20$ by planting no more than $y = 25$ trees.

Solution: nbincdf(5,20,.8) = 0.6167.

The Poisson Random Variable- [https://en.wikipedia.org/wiki/Poisson_distribution] In statistical jargon: Recall that in relation to the binomial random variable we conduct *n* trials, and that each trial can be considered to be a success or a failure. Suppose further that (A1) these trials are mutually *independent*, and that (A2) the probability of success in any given trial is *p*. Now we will address the situation where *n* is very large and where *p* is very small. Let $\lambda = np$. We will now assume that we do not know *n* or *p*, but that we <u>do</u> know $\lambda = np$. In fact, the mean of the bino(*n*,*p*) random variable is exactly $\lambda = np$. Let *Y* denote the total number of successes. Because we do not know *n*, we will assume that it can approach infinity. So, $S_Y = \{0,1,2,\cdots,\infty\}$. Then *Y* has a poisson pdf with parameter *λ*.

In the jargon of random variables: Let Let $\mathbf{X} = (X_1, X_2, \cdots, X_\infty)$ be an ordered ∞-*tuple* of *iid Ber*(*p*) random variables. Now we will address the situation where *n* is very large and where *p* is very small. Let $\lambda = np$. We will now assume that we <u>do not know</u> *n* or *p*, but that we <u>do</u> know $\lambda = np$. Then $Y = \sum_{k=1}^{n} X_k$ is a *poiss*(*λ*) random variable. The pdf for *Y* is:

$f_Y(y) = \frac{e^{-\lambda}\lambda^y}{y!}$ on $S_Y = \{0,1,2,\cdots,\infty\}$.

[Related *Matlab* commands: poisspdf, poisscdf, possinv, poissrnd]

Example Suppose that, on the average, the TSA at a given checkpoint will search 2.5 travelers per hour.

(a) Find the probability that no more than 5 travelers will be searched in an hour.
<u>*Solution*</u>: poisscdf(5,2.5) = 0.9580

(b) Find the probability that no more than 5 travelers will be searched in the next two hours.
<u>*Solution*</u>: poisscdf(5 , 2.5*2) = 0.616

(c) Suppose that while waiting for your friend to arrive, you have noticed that the TSA has searched 7 persons in the last hour. Assuming that the TSA's claimed search rate is 2.5 persons per hour, compute the probability that TSA would search 7 or more persons in an hour.
<u>*Solution*</u>: 1-poisscdf(6,2.5) = 0.0142.

The Hypergeometric Random Variable- [https://en.wikipedia.org/wiki/Hypergeometric_distribution] In statistical jargon: Suppose we have a collection of N objects, wherein K are classified as successes, and $N-K$ are classified as failures. We will draw n objects at random from this collection *without replacement***. The number of successes that we draw, Y, has a hypergeometric pdf with parameters (N,K,n).

In the jargon of random variables: Let Let $\mathbf{X}=(X_1, X_2, \cdots, X_n)$ be an ordered *n-tuple* of Ber(p_k) random variables. In words, X_k is the act of recording a success or failure on the k^{th} draw. Let $Y=\sum_{k=1}^{n} X_k$. Were we to draw objects *with* replacement, then it would be clear that on any given draw, the probability of a success would be simple $p=K/N$. In this case, $\mathbf{X}=(X_1, X_2, \cdots, X_n)$ would be a collection of (A1) mutually independent and (A2) identically distributed Ber(p) random variables. And so Y would be a bino(n,p) random variable. However, we are now drawing *without* replacement. After a little thought, it should be clear that $\{X_k\}_{k=1}^{n}$ are *not mutually independent*. Let's now add some rigor to this claim. Recall that the random variables X_1 and X_2 will be independent only if $\Pr[X_2=1 \mid X_1=1]=\Pr[X_2=1]$. To show that this condition does not hold, note that compute $\Pr[X_1=1]=K/N \overset{\Delta}{=} p$.

Now we will compute $\Pr[X_2=1]$. This is where a clear understanding of events is important. Specifically: $[X_2=1]=[X_1=0 \cap X_2=1] \cup [X_1=1 \cap X_2=1]$. Since these two events are mutually exclusive, we have

$$\Pr[X_2=1]=\Pr[X_1=0 \cap X_2=1]+\Pr[X_1=1 \cap X_2=1].$$

This can be expressed in terms of conditional probabilities as

$$\Pr[X_2=1]=\Pr[\,X_2=1 \mid X_1=0\,] \bullet \Pr[X_1=0] \quad + \quad \Pr[\,X_2=1 \mid X_1=1\,] \bullet \Pr[X_1=1] .$$

We have already noted that: $\Pr[X_1=1]=K/N \overset{\Delta}{=} p$, and so $\Pr[X_1=0]=(N-K)/N$.

It should be equally clear that: $\Pr[\,X_2=1 \mid X_1=0]=K/(N-1)$ and $\Pr[\,X_2=1 \mid X_1=1]=(K-1)/(N-1)$.

Hence, $\Pr[X_2=1]=\left(\dfrac{K}{N-1}\right)\left(\dfrac{N-K}{N}\right)+\left(\dfrac{K-1}{N-1}\right)\left(\dfrac{K}{N}\right)=\dfrac{K}{N} \overset{\Delta}{=} p$.

And so we see that X_2, like X_1, is a *Ber*(p) random variable. [In fact, one can show that every X_k is a *Ber*(p) random variable! In words, $\{X_k\}_{k=1}^{n}$ are *identically distributed Ber*(p) random variables.

Now recall that only if $\Pr[X_2=1 \mid X_1=1]=\Pr[X_2=1]$ will X_1 and X_2 be independent. From the above, we have $\Pr[\,X_2=1 \mid X_1=1]=(K-1)/(N-1)$ and $\Pr[X_2=1]=K/N$. It follows that X_1 and X_2 are not independent.

In conclusion, the collection $\{X_k\}_{k=1}^{n}$ includes *identically distributed Ber*(p) random variables, but they are not mutually independent (as they would be in the case of sampling with replacement). The random variable Y has a hypergeometric(N,K,n) pdf:

$$f_Y(y)=\binom{K}{y}\binom{N-K}{n-y}\Big/\binom{N}{n} \quad \text{on} \quad S_Y=\{\max\{0,n-(N-K)\},\cdots,\min\{K,n\}\}.$$

While the sample space may appear a bit weird, think about it for a moment. If you are drawing *n* objects and $n > N - K$, then you must draw at least $n - (N - K)$ successes.

[Related *Matlab* commands: hygepdf, hygecdf, hygeinv, hygernd]

Example Suppose that one of your favorite CDs has 10 tracks, 3 of which you really, really like. You have just put your CD player into the 'shuffle' mode, where it will randomly select tracks to play. You need to go to class after the 5^{th} track is played.

(a) Find the probability that you will hear exactly 2 of the 3 tracks that you like so much?

Solution: $f_Y(2) = \binom{3}{2}\binom{7}{3}/\binom{10}{5}$ = >> hygepdf(2,10,3,5) = 0.4167

(b) What is the probability that you won't hear any of your best tracks?

Solution: $f_Y(0) = \binom{3}{0}\binom{7}{5}/\binom{10}{5}$ = >> hygepdf(0,10,3,5) = 0.0833

(b) What is the probability that you will hear at least of your best tracks?

Solution: 1-hygecdf(1,10,3,5) = 0.5000