# Lecture 6

# Special Integer-Valued Random Variables

Here we collect a few special discrete random variables that we will run into quite often. All of these random variables assign integer values to the random outcomes. Other varieties will come later.

Recall that a discrete random variable, $X$, is a real-valued function whose domain is the sample space, $S$, and the range, say $\Delta$, is a countable set, such that for each $a \in \Delta$, the resulting partitioning subset of $S$, namely $\{X = a\}$, is an event whose probability is denoted as

$$f(a) \ := \ \mathbb{P}(X = a), \qquad \text{for } a \in \Delta.$$

This function $f$ is called the density of the random variable $X$. Symbolically,

$$S \xrightarrow{X} \Delta \xrightarrow{f} [0,1]; \quad where \quad f(a) = \mathbb{P}(X = a); \quad \text{for each } a \in \Delta.$$

It is understood that $f(a) = 0$ for any $a \notin \Delta$. Other notations for the density of

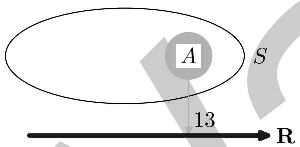


Figure 6.1: An Event, $A = \{X = 13\}$.

a random variable, that we may use in this book, are $f_X(a)$ or $p(a)$ or $p_X(a)$. A cartoon of the event $A = \{X = 13\}$ is shown in Figure 6.1, i.e., $A = \{\omega \in S : X(\omega) = 13\}$. Now we present a few commonly encountered discrete "name-brand" random variables along with their densities, which arise often in real life and are discussed throughout the remainder of the book.





## 6.1 Some Famous Discrete Name Brands

The first example deals with counting the number of successes that we have already seen earlier .

**Example - 6.1.1** - **(Binomial random variables)** We toss a coin twice. The sample space is $S = \{HH, HT, TH, TT\}$. Let $X$ be the number of heads in the two tosses. Now $\Delta = \{0, 1, 2\}$ with $\mathbb{P}(X = 0) = (1-r)^2$, $\mathbb{P}(X = 1) = 2r(1-r)$ and $\mathbb{P}(X = 2) = r^2$, where $r$ is the probability of heads on a single toss. $X$ is called a binomial random variable and we denote it by $X \sim B(2, r)$. The reader should be abel to guess by now what $X \sim B(3, r)$ stands for. In general, if $X$ represents the number of heads in $n$ tosses of a coin, then $X$ has the following probability density:

$$\boxed{\mathbb{P}(X = k) = \binom{n}{k} r^k (1-r)^{n-k}; \qquad k \in \Delta = \{0, 1, 2, \cdots, n\}.} \tag{1.1}$$

A quick reader may recall Example 4.2.3 that proves this. Of course, it is denoted by $X \sim B(n, r)$. The number $r$ (which is the probability of an $H$ on a single toss) represents the type of coin, and is called a parameter of the density. The top two stick graphs in Figure 6.2 show two binomial densities. Unfortunately, there is no

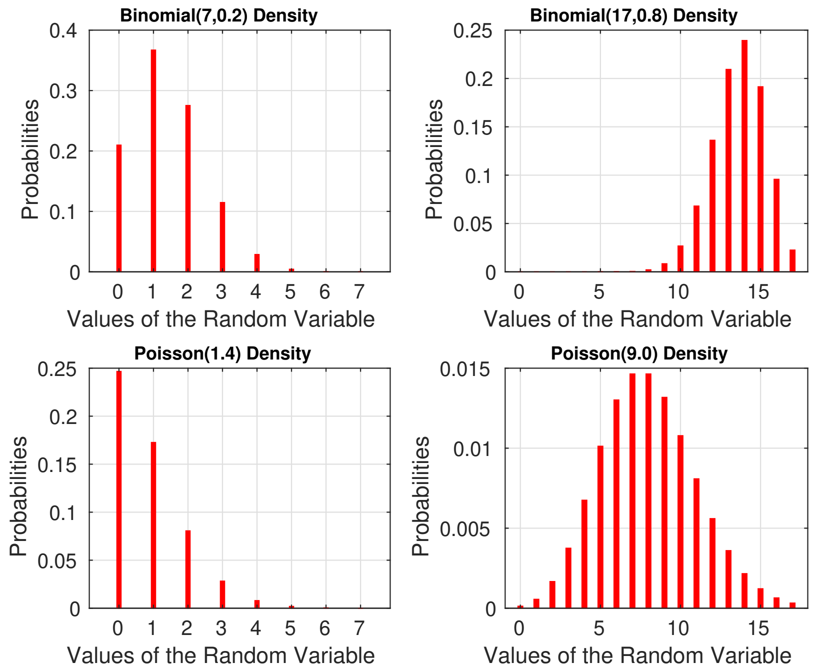


Figure 6.2: $B(7, 0.2)$, $B(17, 0.8)$, $Poisson(1.4)$ and $Poisson(9.0)$ Densities.

closed form formula for the cdf of a binomial rv,

$$F(t) := 0, \quad F(t) := \sum_{k=0}^{m} \binom{n}{k} r^k (1-r)^{n-k}, \quad m \le t < m+1, \qquad m = 0, 1, 2, \cdots, n-1,$$

and $F(t) = 1$ for all $t \ge n$. Modern computational software platforms, such as `Matlab`, `R`, have builtin functions for the binomial cdfs.

**Example - 6.1.2** - **(Poisson random variables)** A random variable $X \in \Delta := \{0, 1, 2, \cdots\}$, for which

$$\boxed{\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}; \qquad k \in \Delta = \{0, 1, 2, \cdots\},} \tag{1.2}$$

where $\lambda > 0$ is a fixed number, is called a Poisson random variable with parameter $\lambda$. We denote this by $X \sim Poisson(\lambda)$. The bottom two stick graphs in Figure 6.2 show two Poisson densities. Poisson distribution is often used as a model to represent the number of telephone calls arriving in a telephone exchange or to model the number of particles emitted by a radioactive substance. Unfortunately, again, there is no closed form expression for the cdf of a Poisson rv,

$$F(t) := 0, \quad t < 0, \qquad F(t) := e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!}, \quad n \le t < n+1, \quad n = 0, 1, 2, \cdots.$$

Modern computing platforms, such as `Matlab`, `R`, have builtin functions for most well known cdfs. The symbolism $X \sim Poisson(\lambda)$ just means that the density of $X$ is Poisson with parameter $\lambda$, as stated in (1.2).

**Example - 6.1.3** - **(Hypergeometric random variables)** In our earlier lectures we have already encountered this variety. Consider an urn containing $G$ number of good apples and $B$ number of bad apples. We reach in and blindly pick $n$ apples all at once. Let $X$ be the number of good applies in our draw. The density of this random variable is

$$\mathbb{P}(X = k) = \frac{\binom{G}{k}\binom{B}{n-k}}{\binom{G+B}{n}}, \qquad k = 0, 1, 2, \cdots, \min\{G, n\}. \tag{1.3}$$

We denote this by $X \sim Hypergeom(G, B, n)$. The shapes of its densities are similar to those of binomial densities. Just like the binomial and Poisson random variables, we are unable to give a general closed form formula for the cdf $F(t)$.

**Example - 6.1.4** - **(Geometric random variables)** If you keep tossing a coin until you get the first $H$, and let $X$ be the number of $T$'s observed before observing the first $H$, then collecting the possible values of $X$, we get $\Delta = \{0, 1, 2, \cdots\}$. The event $\{X = k\}$ will occur only when we observe $k$ successive $T$'s and then an $H$. Since $p$ is the probability of an $H$ on a toss,

$$\boxed{\mathbb{P}(X = k) = p(1-p)^k, \qquad k \in \Delta = \{0, 1, 2, \cdots\}.} \tag{1.4}$$





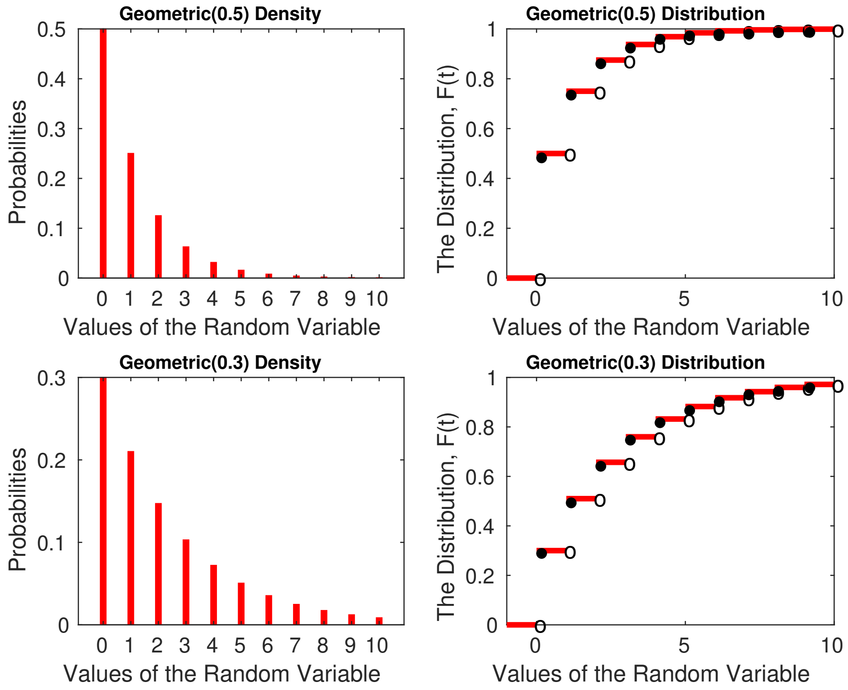


Figure 6.3: Two Geometric Densities and their Distributions.

We use the notation $X \sim Geometric(p)$ to denote this random variable and call it a geometric random variable. Unlike the binomial, Poisson and hypergeometric random variables, the cdf, $F(t) = \mathbb{P}(X \le t)$, of a geometric random variable can easily be written in a closed form. For instance, For $t < 0$, we trivially see that $F(t) = 0$. For $t \in [0, 1)$, $F(t) = \mathbb{P}(X \le t) = \mathbb{P}(X = 0) = p$, and so on. In general,

$$F(t) = \begin{cases} 0 & \text{when } t < 0, \\ \sum_{k=0}^{n-1} p(1-p)^k & \text{when } t \in [n-1, n), \ n = 1, 2, \cdots. \end{cases}$$

Mathematically, when $t \in [n-1, n)$ for $n = 1, 2, \cdots$,

$$F(t) = \mathbb{P}(X \le t) = \sum_{k=0}^{n-1} p(1-p)^k = p \frac{1 - (1-p)^n}{1 - (1-p)} = 1 - (1-p)^n.$$

Figure 6.3 shows two geometric densities along with their corresponding distributions. The symbolism $X \sim Geometric(p)$ means that the density of $X$ is as given in (1.4). Get used to this notation since we will use it through out in our future developments.

**Example - 6.1.5** - **(Negative binomial random variables)** Keep tossing a coin until you get the $n$-th $H$. Let $X$ be the number of $T$'s observed before observing the $n$-th $H$. Collecting the possible values of $X$, we get $\Delta = \{0, 1, 2, \cdots\}$. The event $\{X = k\}$ will occur if and only if when over the resulting $n + k$ spots, the last one is an $H$ and over the first $n - 1 + k$ spots we have $n - 1$ number of $H$'s and $k$ number of $T$'s. Since $p$ is the probability of an $H$ on a toss,

$$\boxed{\mathbb{P}(X = k) \;=\; \binom{n-1+k}{n-1} p^n (1-p)^k, \qquad k \in \Delta = \{0, 1, 2, \cdots\}.} \tag{1.5}$$

We use the notation $X \sim N.B(n, p)$ to denote this random variable and call it a negative binomial random variable. The cdf is not easy to write in a closed form.

## 6.2 Useful Series

For future reference purposes here we collect a few results from Calculus. By induction (or several other ways) we can verify the following arithmetic series,

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}, \quad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}.$$

Take for granted $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$, see also Exercise 6.3.20. Here are a few more series.

$$\begin{aligned}
\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} &= (x+y)^n, \quad \text{(Binomial series)},\\
\sum_{k=0}^{m} \binom{n_1}{k}\binom{n_2}{m-k} &= \binom{n_1+n_2}{m}, \quad \text{(Hypergeometric series)},\\
\sum_{k=0}^{\infty} \frac{x^k}{k!} &= e^x, \quad \text{(Exponential series)},\\
\sum_{k=0}^{\infty} r^k &= \frac{1}{1-r}, \quad \text{(Geometric series)}, \ -1 < r < 1,\\
\sum_{k=0}^{\infty} \binom{n+k-1}{n-1} r^k &= \frac{1}{(1-r)^n}, \quad \text{(Negative binomial series)}, \ -1 < r < 1.
\end{aligned}$$

To verify the hypergeometric series, simply compare the coefficient of $x^m$ on both sides for $(1+x)^{n_1+n_2} = (1+x)^{n_1}(1+x)^{n_2}$. By differentiating the binomial (and other) series we can get more results. For instance, if we differentiate the binomial series with respect to $x$ and then afterwards let $y = 1 - x$, we get

$$\sum_{k=1}^{n} k \binom{n}{k} x^{k-1}(1-x)^{n-k} = n(x+1-x)^{n-1} = n.$$





Twice differentiate the binomial series w.r.t $x$ (or factor switch, cf. Exercise 2.3.13) and then let $y = 1 - x$, to get

$$\sum_{k=2}^{n} k(k-1)\binom{n}{k} x^{k-2}(1-x)^{n-k} = n(n-1)(x+1-x)^{n-1} = n(n-1).$$

Multiplying by $x$ and $x^2$, respectively, on both sides of the last two series we get

$$\sum_{k=1}^{n} k\binom{n}{k} x^k (1-x)^{n-k} \;=\; xn, \quad \sum_{k=2}^{n} k(k-1)\binom{n}{k} x^k(1-x)^{n-k} \;=\; x^2 n(n-1).$$

A similar argument applied to the exponential and the geometric series gives

$$\sum_{k=1}^{\infty} k\frac{x^k}{k!} = xe^x, \qquad \sum_{k=1}^{\infty} kr^k = \frac{r}{(1-r)^2}, \quad -1 < r < 1.$$

Repeated differentiation of geometric series gives the negative binomial series.

**Example - 6.2.1** - Since $2 + 4 + \cdots + (2n) = \sum_{i=1}^{n}(2i) = 2\sum_{i=1}^{n} i = 2\frac{n(n+1)}{2} = n(n+1)$, and $1 + 2 + 3 + 4 + \cdots + (2n) = \frac{(2n)(2n+1)}{2}$, subtracting the first sum from the second sum gives that $1 + 3 + 5 + \cdots + (2n-1) = n(2n+1) - n(n+1) = n^2$.

To see an example of the binomial series in action, take $x = y = 1$ in it to get $\sum_{k=0}^{n}\binom{n}{k} = (1+1)^n = 2^n$. Other choices for $x, y$ give more varieties.

If you take $n_1 = n_2 = m = n$ in the Hypergeometric series you get $\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}$. The reader may now practive with the above useful series to generate many more interesting formulae.

## 6.3 Exercises

**Exercise - 6.3.1** - *At the "dead-man's-curve" on interstate I-90 in Cleveland, accidents occur at the rate of $\lambda = 1$ per month and the number of accident is a Poisson random variable. What is the chance of observing five or more accidents in one month?*

**Exercise - 6.3.2** - *When $X \sim B(n, \frac{1}{2})$, find the probabilities of the following events. (i) $\{X$ is even $\}$, (ii) $\{X$ is odd $\}$.*

**Exercise - 6.3.3** - *When $X \sim Poisson(\lambda)$, find $\mathbb{P}(X$ is even$)$. Note anything peculiar about your answer and explain its cause.*

**Exercise - 6.3.4** - *Find $\mathbb{P}(X$ is even $)$ when $X \sim Geometric(p)$.*

**Exercise - 6.3.5** - *Let $X$ be a random variable that represents the number of tails before observing the 13-th head in repeated tosses of a coin. Assume the probability of observing a head on a single toss is $p$. Show that the density of $X$ is*

$$\mathbb{P}(X = k) = \binom{k+13-1}{13-1} p^{13}(1-p)^k, \qquad k = 0, 1, 2, \cdots.$$

*In other words, show that $X \sim NB(13, p)$.*