Math 321 – Chapter 3 – Random Variables and Probability Distributions
(draft version 2019/04/17-09:20:44)

Contents

1   Random Variables

When you randomly select an individual from a population and measure some trait of that individual, be it categorical (e.g. color, quality level, name, etc.) or quantitative (e.g. mass, height, volume, etc.), you are evaluating a random variable. In this sense a random variable X is quite different from the variable x that we are used to from algebra and calculus. A random variable is a bit more like a function whose input is the outcome from a random experiment, and it outputs the specific observable that we are interested in. Informally, we can think of the random variable looking at the outcome of the experiment, and then and only then, taking on a specific value.

$$X(\text{ outcome of random experiment }) = x$$

This is in a function sense similar to what you are familiar with from algebra and calculus:

$$f(x) = y.$$

Consider the example of rolling a 6-sided die. Let our random variable X be the number that shows on the upper face of the die. We can think of the random variable as a function that takes the actual physical die roll as input and outputs the number of dots on the upper face. The random variable works like this:

$$X(\text{ completed die roll }) = \text{number of dots on upper die face}$$

$$X(\;\boxed{\cdot\cdot\cdot}\;) = 3$$

Normally, we sort of ignore the actual physical outcome of the experiment and just write

$$X = 3$$

to represent the event that the die roll shows 3 on the upper face. We could just say: "the roll is 3" or "the die is 3" or similar.

Generally, As mentioned in Chapter 1, upper-case letters are used to represent the (unevaluated) random variable. Common letters are X, Y, T, Z. A lower-case letter is used to represent a specific value of a random variable, e.g. x, y, t, z. Sometimes we will write something like $X = x$. The capital X on the left represents the (unevaluated) random variable, and the lower-case x on the right indicates an actual numerical value that X can take on, even if it is technically an unspecified numerical value. We could have $x = 3$ or $x = 4$, etc. The x will act similarly to our typical variable form algebra and calculus, i.e. we can treat it as the independent variable for a function (more about this later).

$$X = x \quad \text{means the random variable } X \text{ takes on value } x$$

Now we want to investigate the probability of random variable X taking on specific values, or event the probability of taking on some range of values.

2

2   Probability Distributions

Let's start by considering our die-rolling example. Random variable X is defined as the number of dots on the upper face after a die roll. Now we have a sample space of X-values: $S = \{1, 2, 3, 4, 5, 6\}$. Rather than asking about $P(\text{die roll is } x)$, we can investigate $P(X = x)$. So events can now be about the values a random variable will take. Recall that x is a actually an algebraic variable, and X is the random variable. In this sense $P(X = x)$ is a standard algebraic function of the independent variable x. This function is called the **probability distribution** of random variable X, and is denoted $f_X(x)$. Sometimes it is just denoted $f(x)$ when the underlying random variable is understood, or $p_X(x)$ or $p(x)$ for "probability of x." Often we just caal it the *distribution of X* or the *distribution function.*

2.1   Discrete distributions: probability mass functions

Now it is important here that we are working with a discrete random variable. For a discrete random variable, the probability distribution is usually called the **probability mass function** (usually abbreviated as pmf).

$$f_X(x) = P(X = x) \qquad \text{probability mass function for R.V. } X$$

From now on, we generally won't worry about whether a random variable is categorical or quantitative. Usually we will encode all data with numbers. The outcome of a die roll can even be thought of as a categorical variable. The labels of the sides of a die are really arbitrary and do not need to be dots.

Consider the example of flipping a coin. It is common to use $X = 0$ for tails, and $X = 1$ for heads.

$$X(\text{🪙}) = 0$$

$$X(\text{🪙}) = 1$$

Here is the probability mass function for X if the coin is fair:

$$f_X(x) = \begin{cases} 0.5 & \text{if } x = 0 \\ 0.5 & \text{if } x = 1 \end{cases}$$

If we want to allowed the coin to be biased to a particular side, say with $P(H) = p$, and let p be a parameter, then

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

The distribution $f_X(x)$ is called a probability mass function, and we can think of the probability of an outcome being its "mass."

Example: consider a fair 6-sided die that has painted side that are not marked with numbers of dots. Assume it has 3 green sides, 2 blue sides, and 1 red side. We will define random

variable $X = 1$ for green, $X = 2$ for blue, and $X = 3$ for red. Now we can write the pmf as

$$f_X(x) = \begin{cases} \frac{3}{6} & \text{if } x = 1 \\ \frac{2}{6} & \text{if } x = 2 \\ \frac{1}{6} & \text{if } x = 3 \end{cases}$$

A discrete random variable that takes on a finite list of values is often called **uniform**. If X can take on values $\{a, a+1, a+2, \ldots, b\}$ and is uniform (all values equally likely), then X is called a **uniform random variable on** $\{a, b\}$.

2.2   Continuous distributions: probability density functions

For continuous random variables, the total number of allowable values is *uncountably infinite*. In this case the probability distribution is called the **probability density function** (usually abbreviated as pdf). The value of the pdf at a point is no longer a true probability, it is a probability density. To get a probability, one must integrate the pdf. The function $f_X(x)$ is a probability density function for random variable X if

$$P(a < X < b) = \int_a^b f_X(x)dx.$$

We will often say it this way: "f is the distribution of X" or "f is the pdf of X."

For a continuous random variable X, $P(X = x) = 0$ for all individual x-values. The probability of each individual value is zero! There is a bit of a paradox here in that probability zero does not mean impossible! Don't get bogged down with this. Just think of it as a mathematical convenience. The integral over a single point is always zero.

Example: $f(x) = kx$ for $0 \leq x \leq 5$. Find k. We must find the k value that makes the total integral equal to 1 (law of total probability). $\int_0^5 kxdx = 1$ gives $\left[\frac{k}{2}x^2\right]_0^5 = k\frac{25}{2} = 1$ thus $k = \frac{2}{25}$.

Example: This requires you to recall how to compute improper integrals. $f(x) = kx^{-3}$ for $x \geq 1$. Let's find k. $1 = \int_1^\infty kx^{-3}dx = -\frac{k}{2}x^{-2}\Big|_1^\infty = -\frac{k}{2 \cdot \infty^2} + \frac{k}{2} = \frac{k}{2}$ thus $k = 2$.

2.3   Cumulative distributions

For any random variable X, whether discrete or continuous, $F_X(x) = P(X \leq x)$ is called the cumulative distribution function, or cdf.

Properties of the cdf:

- $F(x)$ goes from 0 to 1, this means that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

- The domain of F is $(-\infty, \infty)$.

- If F has any jump discontinuities, the size of the jump is the probability of the x-value where the jump occurs. This means that $P(X = x) = F(x) - \lim_{k \to x-} F(k)$.

- $P(a < X \le b) = F(b) - F(a)$.

2.3.1 Discrete CDF

Calculating the probability of a range of values with a discrete cdf can be tricky depending on the precise inequalities involved. Here are the four variations.

If discrete random variable X takes on values $\{a, a+1, a+2, \ldots, b-1, b\}$ then:

$$
\begin{aligned}
P(a < X \le b) &= F(b) - F(a) \\
P(a \le X \le b) &= P(a - 1 < X \le b) \\
&= F(b) - F(a - 1) \\
P(a < X < b) &= P(a < X \le b - 1) \\
&= F(b - 1) - F(a) \\
P(a \le X < b) &= P(a - 1 < X \le b - 1) \\
&= F(b - 1) - F(a - 1)
\end{aligned}
$$

More generally, replace $a - 1$ and $b - 1$ in the above formulas for the "next lowest possible x-value".

Example:

pmf in table format:

x	1	5	10	20
$f(x)$	0.9	0.05	0.04	0.01

pmf in piecewise format:

$$
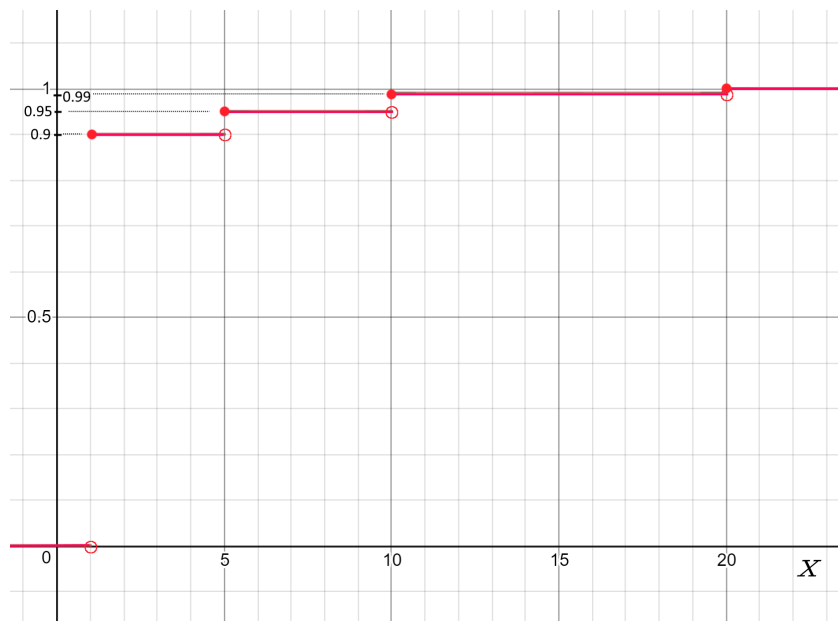f(x) = \begin{cases}
0.9 & \text{if } x = 1 \\
0.05 & \text{if } x = 5 \\
0.04 & \text{if } x = 10 \\
0.01 & \text{if } x = 20 \\
0 & \text{for all other } x \text{ values}
\end{cases}
$$

cdf in piecewise format:

$$
f(x) = \begin{cases}
0 & \text{for } x < 1 \\
0.9 & \text{if } 1 \le x < 5 \\
0.95 & \text{if } 5 \le x < 10 \\
0.99 & \text{if } 10 \le x < 20 \\
1 & \text{if } 20 \le x
\end{cases}
$$

See the graph of the cdf below:

Calculate the probabilities:

$P(1 < X \le 10) = F(10) - F(1)$. This is the probability of the values $x = 5, 10$ only.

$P(1 < X < 10) = P(1 < X \le 5) = F(5) - F(1)$

$P(1 \le X < 10) = P(X \le 5) = F(5)$

$P(5 \le X \le 10) = P(1 < X \le 5) = F(5) - F(1)$

$P(5 < X < 10) = P(5 < X \le 5) = F(5) - F(5) = 0$

2.3.2   Continuous CDF

The cdf for a continuous random variable X is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du.$$

In many cases, the lower limit of the integral need not be $-\infty$.

Example: Consider pdf: $f(x) = \frac{1}{2}$ for $0 \le x \le 2$. The cdf is $F(x) = \int_0^x \frac{1}{2}du = \left[\frac{u}{2}\right]_0^x = \frac{x}{2}$. We do not need to integrate from $-\infty$ to 0 because the pdf is zero there and thus will not contribute to the integral.

2.4   Percentiles of distributions

Percentiles/quantiles for continuous distributions:

The $(100p)^{th}$ percentile for continuous RV X with pdf $f(x)$ is notated by $\tilde{x}_p$ and is given by

$$p = \int_{-\infty}^{\tilde{x}_p} f(x)dx = F(\tilde{x}_p)$$

6

Thus solving for the inverse function of F we get $F^{-1}(p) = \tilde{x}_p$.

In particular, the median $\tilde{\mu}$ is the 50th percentile and is given by $\tilde{\mu} = F^{-1}(0.5)$.

Percentiles/quantiles for discrete distributions:

Let us assume we know the cdf $F(x)$ and that $x = x_1, x_2, \ldots, x_k$ with probabilities $p_1, p_2, \ldots, p_k$ (note that $\sum_i p_i = 1$ by the law of total probability). Then there are k total x-values, and x_i is the $F(x_i)100^{th}$ percentile.

Example: Using the tabular example from above, we can see that $x = 1$ is the 90th percentile, $x = 5$ is the 95th percentile, $x = 10$ is the 99th percentile, and $x = 20$ is the 100th percentile.

3 Expected value

Discrete RV:

$$E(X) = \sum_{\text{all } x} x f_X(x)$$

Continuous RV:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Usually we use the letter μ for expected value: $\mu = E(X)$.

Discrete Example: Use the table in the previous example. $E(X) = \sum_{\text{all } x} x f_X(x) = (1)(0.9) + (5)(0.05) + (10)(0.04) + (20)(0.01) = 1.75$. So if we think of X as a gambling payout, then you should expect to win \$1.75. This can be calculated in R as follows:
```
> x=c(1,5,10,20);
  p=c(0.9,0.05,0.04,0.01);
  EX=sum(x*p)
```

Continuous example: Consider pdf $f(x) = \frac{2}{25}x$ for $x \in [0,5]$. The expected value is $E(X) = \int_0^5 \frac{2}{25} x^2 dx = \left[\frac{2}{75} x^3\right]_0^5 = \frac{2}{75}(125) - \frac{2}{75}(0) = \frac{10}{3}$.

Example: $f(x) = 2x^{-3}$ for $x \geq 1$. $E(X) = \int_1^{\infty} x \cdot 2x^{-3} dx = \int_1^{\infty} 2x^{-2} dx = -x^{-1}|_1^{\infty} = 1$.

3.1 Expected value of a function of a random variable

If we have a function of a random variable $h(X)$, then this is a new random variable, say we call it $Y = h(X)$. If we want to study the random behavior of Y, it is completely determined by X. In particular its expected value is $E(Y) = E(h(X))$.

$$\text{Discrete:} \quad E(h(X)) = \sum_x h(x) f(x)$$

$$\text{Continuous:} \quad E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x)$$

7

In particular, it is important to be able to calculate $E(X^2)$ as we'll see below.

Expectation for a linear function of a random variable is as follows:
$$E(aX + b) = aE(X) + b$$

Note that we can now take the expectation of sums of random variables:
$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

Example: Consider pmfs given by the tables below.

x	1	2	3	4
$f_X(x)$	0.4	0.2	0.1	0.3

y	10	20	30	40
$f_Y(y)$	0.1	0.1	0.2	0.6

Calculate $E(Y - 10X)$. $E(Y - 10X) = E(Y) - 10 \cdot E(X) = 33 - 10 \cdot 2.3 = 10$.

4 Variance

$$\begin{aligned} \text{Var}(X) &= E\left((X - E(X))^2\right) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Usually we use the letter σ^2 for variance: $\sigma^2 = \text{Var}(X)$.

Discrete Example: Use the table in the previous example.
$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 \\ &= (1)^2(0.9) + (5)^2(0.05) + (10)^2(0.04) + (20)^2(0.01) - (1.75)^2 \\ &= 7.0875 \end{aligned}$$

This can be done in R as follows:
```
> x=c(1,5,10,20);
  p=c(0.9,0.05,0.04,0.01);
  EX=sum(x*p)
  VarX=sum(x^2*p)-EX^2
```

Continuous example: Consider pdf $f(x) = \frac{2}{25}x$ for $x \in [0, 5]$. The expected value is $E(X) = \frac{10}{3}$ thus we calculate the variance as:

$$\begin{aligned} \text{Var}(X) &= \int_0^5 x^2 f(x)dx - \left(\frac{10}{3}\right)^2 \\ &= \int_0^5 x^2 \frac{2}{25}xdx - \left(\frac{10}{3}\right)^2 \\ &= \left[\frac{2}{100}x^4\right]_0^5 - \left(\frac{10}{3}\right)^2 \\ &= \frac{25}{18}. \end{aligned}$$

Numerical integration can be done in R as follows:
```
> f = function(x){f(x)}
  integrate(f,lower=a,upper=b)$value
```

So here is how we will solve this problem in R:
```
> xf = function(x){x * 2/25*x}
  EX = integrate(xf,lower=0,upper=5)$value
  x2f = function(x){x * 2/25*x}
  EX2 = integrate(x2f,lower=0,upper=5)$value
  VarX = EX2-EX^2
```

Example with infinite variance: $f(x) = 2x^{-3}$ for $x \geq 1$. We already calculated $E(X) = 1$ above. $E(X^2) = \int_1^\infty x^2 \cdot 2x^{-3}dx = \int_1^\infty 2x^{-1}dx = 2\ln x|_1^\infty = 2\ln\infty - 2\ln 1 = 2\ln\infty = \infty$. Thus $Var(X) = E(X^2) - (E(X)^2) = \infty - 1 = \infty$. We can have a probability density function that has infintie mean as well, take $f(x) = \frac{1}{x^2}$ for $x \geq 1$.

4.1 Variance of a function of a random variable

If we have a function of a random variable $h(X)$, then this is a new random variable, say we call it $Y = h(X)$. If we want to study the random behavior of Y, it is completely determined by X. In particular its expected value is $Var(Y) = Var(h(X))$.

$$\text{Var}(h(X)) = E((h(X))^2) - (E(h(X))^2$$

Expectation for a linear function of a random variable is as follows:

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

5 Special discrete distributions

5.1 Bernoulli

Situation/explanation: We perform a single random experiment that has only two outcomes: success and failure. Success and failure can mean almost anything. Success can mean a no vote on a particular election item, or it can mean we find a defective item, or it can mean a die roll has a particular value.

$X = 0$ for failure, $X = 1$ for success, $p =$ probability of success.

$$X \sim \text{Bernoulli}(p)$$
$$X = 0, 1$$
$$f_X(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$$

$$E(X) = p$$
$$\text{Var}(X) = p(1-p)$$

In R:
$$f(x) = P(X = x) = \texttt{dbinom(x,size=1,prob=p)}.$$
$$F(x) = P(X \leq x) = \texttt{pbinom(x,size=1,prob=p)}.$$

5.2 Binomial

Situation/explanation: We perform a fixed number of Bernoulli trials. Count the total number of successes.

X = number of successes out of n Bernoulli trials with p = probability of success.

$$X \sim \text{Bin}(n,p)$$
$$X = 0, 1, 2, 3, \ldots, n$$
$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E(X) = np$$
$$\text{Var}(X) = np(1-p)$$

In R:
$$f(x) = P(X = x) = \texttt{dbinom(x,size=n,prob=p)}. \quad \text{(pmf)}$$
$$F(x) = P(X \leq x) = \texttt{pbinom(x,size=n,prob=p)}. \quad \text{(cdf)}$$
$$\tilde{x}_q = F^{-1}(q) = \texttt{qbinom(q,size=n,prob=p)}. \quad (q100^{th} \text{ percentile, inverse cdf)}$$

Simulate N independent draws from a $\text{Bin}(n,p)$ distribution:
$$\texttt{rbinom(N,size=n,prob=p)}.$$

Example: Consider the experiment of flipping a fair coin 1,000 times. Calculate the probability that there are at least 530 heads but less than 600.

We have that $X \sim \text{Bin}(n = 1000, p = 0.5)$, and want to calculate $P(530 \leq X < 600)$.

$$P(530 \leq X < 600) = P(529 < X \leq 599) = F(599) - F(529)$$

Here is a way to solve this in R:
```
> n=1000;
  p=1/2;
  Xmin=529;
  Xmax=599;
```

10

```
pbinom(Xmax,size=n,prob=p)-pbinom(Xmin,size=n,prob=p)
```

[1] 0.0310116

So there is a 3.1% chance that the number of heads out of 1,000 flips of a fair coin will be greater than or equal to 530 but less than 600.

5.3 Geometric

Situation/explanation: We perform Bernoulli trials until we get our first success. Count the total number of trials (a bunch of failures, and one success).

X = total number of trials required to get a single success with p = probability of success. (the count includes the success also)

Note that there is no fixed total number of trials here. Note: there is a single success, the last trial, the first $x - 1$ trials are all failures.

$$X \sim \text{Geom}(p)$$

$$X = 0, 1, 2, 3, \ldots$$

$$f(x) = p(1-p)^{x-1}$$

$$F(x) = 1 - (1-p)^x$$

$$E(X) = \frac{1}{p}$$

$$\text{Var}(X) = \frac{(1-p)}{p^2}$$

(Note that in R, the definition of X varies from this. In R, X is the number of failures. So when calculating in Rfor the Geometric random variable, we must either subtract 1 from the x values in R for the pmf and cdf, but add 1 to the output from an R quantile.)

In R:

$$f(x) = P(X = x) = \texttt{dgeom}(x\text{-1},\texttt{prob=}p). \quad \text{(pmf)}$$

$$F(x) = P(X \le x) = \texttt{pgeom}(x\text{-1},\texttt{prob=}p). \quad \text{(cdf)}$$

$$\tilde{x}_q = F^{-1}(q) = \texttt{qgeom}(q,\texttt{prob=}p)\text{+1}. \quad (q100^{th} \text{ percentile, inverse cdf})$$

Simulate N independent draws from a $\text{Geom}(p)$ distribution: `rgeom(N,prob=p)+!`.

Example: flip a fair coin until the first H. Let X be the number of tails. This is geometric with $p = \frac{1}{2}$. $P(X = x) = \left(\frac{1}{2}\right)^x$

Example: Let a production line produce a defective item with probability 0.01. Let X be the number of items produced before the first defective item. (Say once a defective item is produced, the machinery must be stopped and recalibrated.) This is geometric with $p = 0.01$. $P(X = x) = 0.01(0.99)^{x-1}$

5.4 Negative binomial

Situation/explanation: We perform Bernoulli trials until we observe the desired number of successes. Count the number of failures.

$X=$ number of failures required to get r successes with $p =$ probability of success

$$X \sim \text{NegBin}(r, p)$$

$$X = 0, 1, 2, 3, \dots$$

$$f_X(x) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x$$

$$E(X) = \mu = \frac{r(1 - p)}{p}$$

$$Var(X) = \frac{r(1 - p)}{p^2}$$

In R:

$$f(x) = P(X = x) = \texttt{dnbinom}(x, \texttt{size=}r, \texttt{prob=}p). \quad \text{(pmf)}$$

$$F(x) = P(X \leq x) = \texttt{pnbinom}(x, \texttt{size=}r, \texttt{prob=}p). \quad \text{(cdf)}$$

$$\tilde{x}_q = F^{-1}(q) = \texttt{qnbinom}(q, \texttt{prob=}p). \quad (q100^{th} \text{ percentile, inverse cdf)}$$

Simulate N independent draws from a NegBinom(r, p) distribution:
$$\texttt{rnbinom}(N, \texttt{size=}r, \texttt{prob=}p).$$

Example: We want to flip a coin until we observe 3 heads. Count the *number of tails*. $X =$ the number of tails, $X = 0, 1, 2, 3, \dots$ (could theoretically go all the way up to infinity). The experiment stops after 3 heads. Note that the last coin flip is going to be the 3rd head always.

Example: Let a production line produce a defective item with probability 0.01. Let X be the number of non-defective items produced once 100 defective ones are produced. Let's say, that after 100 defective items, the machinery will be recalibrated. Now X is geometric with $p = 0.01$. $P(X = x) = \binom{x+100-1}{10-1}(0.01)^{100}(0.99)^x$. Calculate the probability that at least 7,500 non-defective items are produced before recalibration is necessary. $P(X \geq 7500) = 1 - P(X \leq 7499) = 1 - F(7499)$. In R this can be done as $\texttt{1-pnbinom(7499,size=100,prob=0.01)} \approx 99.5\%$.

5.5 Hypergeometric

Situation/explanation: There is a fixed number of outcomes (M successes and $N - M$ failures, for a total of N outcomes to choose from). We randomly select n outcomes (some of them will be successes, and the rest failures). We want to count the number of successes in our sample of size n.

$X=$ number of successes in a sample of size n out of a "bag" of M successes and $N - M$ failures

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad \text{for} \quad \max(0, n - N + M) \leq x \leq \min(n, M)$$

$$E(X) = \mu = n\frac{M}{N}$$

$$Var(X) = n\frac{M}{N}\left(\frac{N-n}{N-1}\right)\left(1 - \frac{M}{N}\right)$$

In R:

$$f(x) = P(X = x) = \texttt{dhyper(x,M,N-M,n)}. \quad \text{(pmf)}$$

$$F(x) = P(X \leq x) = \texttt{phyper(x,M,N-M,n)}. \quad \text{(cdf)}$$

$$\tilde{x}_q = F^{-1}(q) = \texttt{qhyper(q,M,N-M,n)}. \quad (q100^{th} \text{ percentile, inverse cdf})$$

Simulate N_{sim} independent draws from a HyperGeom$(M, N - M, n)$ distribution:
$$\texttt{rhyper}(N_{sim}\texttt{,M,N-M,n}).$$

Example: A box contains 15 balls, 4 of them are red, 11 are blue. We will draw 3 of them. Count the number of red balls in our sample of 3. X = number of red balls in our sample, $0 \leq X \leq 3$.

5.6 Poisson

Situation/explanation: (1) Events arrive over time. Count the number of events in a given time interval. (2) Events are distributed over a spatial region randomly. Count the number of events in a given region.

X = number of events that occur at rate λ. The rate can be thought of as "the number of events per unit time" or more generally, "number of events per unit." Often it is the number of events per unit length or per unit time.

$$X \sim \text{Pois}(\lambda)$$
$$X = 0, 1, 2, \ldots$$
$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$
$$E(X) = \lambda$$
$$Var(X) = \lambda$$

In R:

$$f(x) = P(X = x) = \texttt{dpois(x,lambda=}\lambda\texttt{)}. \quad \text{(pmf)}$$

$$F(x) = P(X \leq x) = \texttt{ppois}(x,\texttt{lambda=}\lambda). \quad \text{(cdf)}$$

$$\tilde{x}_q = F^{-1}(q) = \texttt{qpois}(q,\texttt{lambda=}\lambda). \quad (q100^{th} \text{ percentile, inverse cdf})$$

Simulate N_{sim} independent draws from a Pois(λ) distribution: $\texttt{rpois}(N_{sim},\texttt{lambda=}\lambda)$.

Example: (1) Information requests arrive at a server randomly, the number of requests in an hour will be Poisson. (2) The number of cars that pass through an intersection in an 8 hour period. (3) A manufacturer produces rope. Each rope is 100ft long. The number of imperfections on a 100ft length of rope will be Poisson. (4) A publisher produces pages of text. The number of printing errors on a given page is Poisson.

5.6.1 Scaling the Poisson parameter

The Poisson rate parameter can be scaled to fit the situation, e.g. if a 2 square meter bolt of clothe has on average 3 defects, the rate is 3 defects per 2 square meters of clothe. We can actually scale the parameter λ to fit any situation. If we are interested in a 10 square meter bolt of clothe, then we can scale our rate parameter up: $10/2 = 5$ so that we have 5 2 square meter sections, and each should on average have 3 defects, thus we have a total of 6 defects expected for the 10 sq. m. section.

If the rate is k events per m units, then $\lambda = k$ (if we are interested in number of events in m units).

If we are interested in the number of events in 1 unit, then we can scale, the rate parameter for 1 unit is $\lambda = k/m$.

If we are interested in the number of events in t units, then we can scale, the rate parameter for t units is $\lambda = k/m \cdot t$.

Example: Let's say a book publisher produces pages of text that have on average, 3 defects per 700 pages of text. (Such defects can be paper tears, ink smudges, etc.) If a particular book has 1,200 pages, then we can model the number of defects found in a copy of that book as a Poisson random variable. The rate parameter is $3/700$ defects per page, so we'll scale this up to 1,200 pages: $\lambda = 3/700 \cdot 1200 = \frac{36}{7} \approx 5.14$. So we expect this book to have on average about 5 defects.

Continuing this example, a book with exactly 700 pages will have $\lambda = 3$, abook with 1,400 pages will have $\lambda = 6$, and a book with N pages will have $\lambda = \frac{3}{700} \cdot N$.

6 Special continuous distributions

6.1 Uniform

The *uniform* random variable is the continuous analog of "equally likely".

$$X \sim \mathrm{Unif}(a, b)$$
$$a \leq X \leq b$$
$$f_X(x) = \frac{1}{b-a} \quad \text{for } x \in [a, b]$$
$$F(x) = \frac{x-a}{b-a}$$
$$E(X) = \frac{a+b}{2}$$
$$\mathrm{Var}(X) = \frac{1}{12}(b-a)^2$$

In R:
$$f(x) = \texttt{dunif}(x, \texttt{min=}a, \texttt{max=}a). \quad \text{(pdf, not probability mass)}$$
$$F(x) = P(X \leq x) = \texttt{punif}(x, \texttt{min=}a, \texttt{max=}a). \quad \text{(cdf)}$$

6.2 Exponential

Suppose events happen at random times (or at random locations along a physical length). The length of time between events can be modeled by an *exponential random variable.*

$X =$ wait time until an event.
Rate parameter λ is the number of events per unit time. It can be number of events per unit length as well; it just depends on the context. This is very closely related to the Poisson, as we'll discuss later.

We can refer to X as the "inter-arrival time," "wait time," or "inter-event time."

$$X \sim \mathrm{Exp}(\lambda)$$
$$0 \leq X$$
$$f_X(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$
$$F(x) = 1 - e^{-\lambda x}$$
$$E(X) = \frac{1}{\lambda}$$
$$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$

In R:
$$f(x) = \texttt{dexp}(x, \texttt{rate=}\lambda). \quad \text{(pdf, not probability mass)}$$
$$F(x) = P(X \leq x) = \texttt{pexp}(x, \texttt{rate=}\lambda). \quad \text{(cdf)}$$

$$\tilde{x}_q = F^{-1}(q) = \texttt{qexp}(q, \texttt{rate=}\lambda). \quad (q100^{th} \text{ percentile, inverse cdf})$$

Simulate N_{sim} independent draws from a $\mathrm{Exp}(\lambda)$ distribution: $\texttt{rexp}(N_{sim}, \texttt{rate=}\lambda)$.

6.2.1 Connection between Poisson and exponential

There is a reason that I use the same Greek letter λ for both Poisson and exponential random variables. The Poisson distribution models the number of events in a given interval, and the exponential distribution models the time or space between the events.

Example: Assume that a particular roadway intersection has cars arriving at rate 25 cars every 2 hours. Then the time between successive arrivals of cars is modeled by an exponential random variable $X_{\text{inter-arrival time}} \sim Exp(\lambda = 25)$. If we are interested in the number of cars that pass through the intersection in a given hour, then it is $X_{\text{number of cars in 2 hours}} \sim Pois(\lambda = 25)$. Now if we are interested in a different length of time, we need to scale the Poisson, but the exponential never needs to be scaled in this way. If we are interested in a particular 8 hour work day, then the number of cars that pass through the intersection is Poisson with mean

$$\lambda T = \{25 \ cars\}/\{2 \ hours\} \times \{8 \ hours\} = 100 \ cars$$

$$X_{\text{number of cars in 8 hours}} \sim Pois(\lambda T) = Pois(25/2 \cdot 8) = Pois(100)$$

If we were to calculate with this in R, we would need to use `ppois(x,lambda=100)` for example.

If we are interested in a single hour:

$$X_{\text{number of cars in 1 hour}} \sim Pois(\lambda T) = Pois(25/2 \cdot 1) = Pois(12.5)$$

But in all cases, whether we are interested in a single hour, 8 hours, or 24 hours, etc., the inter-arrival time (or wait time) between successive cars at the intersection is exponential with rate parameter $\lambda = 25/2$.

$$X_{\text{wait time between cars}} \sim Exp(12.5)$$

In R this would be calculated as, for example, $P(X_{wait} \leq x) =$`pexp(x,rate=12.5)`.

6.3 Normal

$$X \sim \text{N}(\mu, \sigma^2)$$
$$-\infty < X < \infty$$
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } x \in \mathbb{R}$$
$$F(x) = \text{Unfortunately, can't be written down in a closed fomula}$$
$$E(X) = \mu$$
$$\text{Var}(X) = \sigma^2$$

In R:
$$f(x) = \texttt{dnorm}(x, \texttt{mean=}\mu, \texttt{sd=}\sigma). \quad \text{(pdf, not probability mass)}$$
$$F(x) = P(X \leq x) = \texttt{pnorm}(x, \texttt{mean=}\mu, \texttt{sd=}\sigma). \quad \text{(cdf)}$$

$$\tilde{x}_q = F^{-1}(q) = \texttt{qnorm}(q, \texttt{mean=}\mu, \texttt{sd=}\sigma). \quad (q100^{th} \text{ percentile, inverse cdf})$$

Simulate N_{sim} independent draws from a $N(\mu, \sigma^2)$ distribution:
$$\texttt{rnorm(}N_{sim}\texttt{,mean=}\mu\texttt{,sd=}\sigma\texttt{)}.$$