CHAPTER 2:

2.1. You need an itinerary to visit 6 cities in succession. The 6 cities are

Boise, ID	Seattle, WA	Denver, CO
Miami, FL	Atlanta, GA	Newark, NJ

For example, one itinerary is

$$\text{Atlanta} \longrightarrow \text{Denver} \longrightarrow \text{Newark} \longrightarrow \text{Boise} \longrightarrow \text{Miami} \longrightarrow \text{Seattle}.$$

Note that the ordering of the cities uniquely determines the itinerary. All 6 cities must be included.
(a) How many possible itineraries are there?
(b) What does it mean when we say "each possible itinerary is equally likely?"
(c) Suppose you pick one itinerary at random from all possible itineraries. What is the probability the itinerary you select has the West coast cities (Boise, Seattle, and Denver) and the East cost cities (Miami, Atlanta, and Newark) grouped together? *Hint:* This means you visit the 3 West coast cities first and then the 3 East coast cities second (or vice versa).

2.2. A construction company hires two engineers to bid on potential jobs: Engineer 1 and Engineer 2.

- Engineer 1 bids on 70% of the jobs; Engineer 2 bids on the remaining jobs.
- Among all jobs on which Engineer 1 bids, an error is made on 2% of the bids.
- Among all jobs on which Engineer 2 bids, an error is made on 4% of the bids.

(a) Define two relevant events A and B using the information above. Interpret each percent above as a probability (two are conditional probabilities). Do not define more than two events or you will be making the problem too hard.
(b) What proportion of all bids will contain an error? Which "law" are you using here?
(c) Suppose a new bid sent to the manager contains an error. Find the probability this bid was prepared by Engineer 1.

2.3. In a recent report from the US Health Resources and Services Administration, the following population level characteristics were presented:

- 1 percent of Americans were infected with HCV
- among those Americans infected with HCV, 3 percent were infected with HIV
- among those Americans infected with HIV, 10 percent were infected with HCV.

(a) Define two relevant events and interpret the three percentages above in terms of probabilities (two are conditional probabilities).
(b) Find the percentage of Americans with HIV.
(c) Are these two diseases independent?

2.4. A manufacturer buys 60 percent of a raw material (e.g., nails) from Supplier 1 and 40 percent from Supplier 2. Two percent of the raw materials from Supplier 1 are defective. Four percent of the raw materials from Supplier 2 are defective.
(a) Define two relevant events A and B using the information above. Interpret each percent above as a probability. Do not define more than two events or you will be making the problem too hard.
(b) A piece of raw material is selected from the production line at random. What is the probability the raw material selected is defective? What "law" are you using here?
(c) If the raw material selected was defective, what is the probability it came from Supplier 1?

2.5. A car repair is either completed on time or it is not completed on time. A repair is either satisfactory or not satisfactory.

- If a repair is completed on time, the probability it is satisfactory is 0.90.
- If a repair is not completed on time, the probability it is satisfactory is 0.95.
- The probability a repair is completed on time is 0.75.

(a) Define two events A and B and write each number above using our notation for probability and conditional probability. Do not define more than two events (aside from complements) or you will make the problem too hard.
(b) What is the probability a repair is completed on time and is satisfactory?
(c) What is the probability a repair is not satisfactory?
(d) A mechanic believes whether a repair is satisfactory has nothing to do with whether the repair is completed on time. Is he correct?

CHAPTER 3:

3.1. At a local plant, the random variable

$$X = \text{number of serious accidents per year}$$

has the following probability mass function (pmf):

x	0	1	2	3	4
$p_X(x)$	0.5	0.2	0.1	0.1	0.1

(a) Explain why X is a discrete random variable (and not continuous).
(b) Find the mean number of serious accidents per year.
(c) Prepare a graph of $F_X(x)$, the cumulative distribution function of X. Label both axes and use appropriate tick marks on each axis. Neatness counts.

3.2. A batch contains 36 bacteria cells. Assume 24 of the cells are capable of replication; the remaining 12 are not. From the batch of 36 cells, a geneticist will sample 3 cells at random and without replacement.
(a) How many samples of 3 cells are possible? The order the cells are selected is not important.

(b) Let X denote the number of cells capable of replication among the 3 chosen by the geneticist. Find the probability mass function (pmf) of X. *Hint:* What are the possible values of X? Is X discrete or continuous?
(c) Prepare a graph of the cumulative distribution function (cdf) of X. Neatness counts. If you could not complete part (b), you can still draw a graph of what you think the cdf of X would look like.

3.3. A web host has 4 independent servers connected in parallel. At least 3 of them must be operational for the web service to be operational.
(a) If individual servers are operational with probability 0.95, what is probability the web service is operational? Answer this question by using the binomial distribution; e.g., let X denote the number of operational servers (out of 4).
(b) For this part, suppose the individual servers have different probabilities of being operational: 0.70, 0.80, 0.99, and 0.99, respectively. What is the probability the web service is operational now? Why can't you use the binomial distribution to answer this part?

3.4. Suppose 90 percent of all batteries from a supplier have acceptable voltages for operation. A random sample of 20 batteries is collected and the batteries are randomly assigned to flashlights. Each flashlight has two batteries. There are 10 flashlights total. For a flashlight to be operational, both batteries in it must have acceptable voltages. Assume each battery operates independently of other batteries.
(a) Calculate the probability a single flashlight is operational.
(b) Among the 10 flashlights, what is the probability that at least 9 will be operational?
(c) Redo parts (a) and (b) under the assumption that a flashlight is operational if at least one of its two batteries has an acceptable voltage. Continue to assume all batteries are independent.

3.5. A blood bank receives donors in succession. Here is the distribution of the blood types in the US population:

O^+	O^-	A^+	A^-	B^+	B^-	AB^+	AB^-
0.38	0.07	0.34	0.06	0.09	0.02	0.03	0.01

Treat each donor visiting the bank as independent.
(a) What is the probability the first O^+ blood type donor will be seen among the first 4 donors who visit the bank?
(b) The bank will remain open until it receives 5 donors who are AB^+ or AB^-. What probability distribution describes the number of donors that will be seen?
(c) In part (b), what is the mean number of donors that will be seen in total? If you cannot remember the correct formula, use your intuition and explain.

3.6. On February 27, 2013, the City Council of Cincinnati (OH) passed an ordinance requiring photoelectric smoke detectors in all rental properties. However, over 5 years later, the city's fire department representatives are concerned not all properties are adhering to the ordinance. Suppose the population proportion of rental properties in Cincinnati having photoelectric smoke detectors is 0.80 (i.e., 80 percent).

(a) If a sample of 6 rental properties is selected at random, what is the probability at least 5 have photoelectric smoke detectors installed?
(b) What three Bernoulli trial assumptions did you make in part (a)?
(c) Suppose rental properties were inspected until the first one without photoelectric smoke detectors was found. Under the assumptions you outlined in part (b), what is the distribution of the number of rental properties that would be inspected?

3.7. The number of automobiles entering a remote mountain tunnel has a Poisson distribution with mean $\lambda = 10$ cars per hour.
(a) Find the probability there are at most 2 cars that enter the tunnel in a given hour.
(b) (Chapter 4) In any given hour, what is the probability we would have to wait at least 30 minutes for the first car to enter the tunnel? Note that 30 minutes is 1/2 of an hour.

3.8. The South Carolina Department of Revenue estimates 10 percent of all individual state income tax forms filed during 2016 will contain "serious errors."
(a) Conceptualizing each tax form filed as a "trial," state the 3 Bernoulli trial assumptions in the context of this problem. Assume these hold for the parts below.
(b) If an auditor processes 20 tax forms, calculate the probability at least 2 will contain serious errors.
(c) An auditor processes tax forms until he finds the first one with serious errors. What is the probability he will process at most 3 forms?

3.9. A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates the defective rate of the device is 3%.
(a) An inspector randomly picks 10 devices and tests them. What is the probability there are 1 or 2 defective devices? *Hint:* Use the binomial distribution.
(b) Your calculation in part (a) requires three assumptions. Describe what these assumptions are in this problem.

3.10. Automobiles arrive at a vehicle equipment inspection station according to a Poisson process with mean $\lambda = 4$ per hour.
(a) What is the probability at most 2 vehicles will arrive in a given hour?
(b) (Chapter 4) Let T denote the time until the first vehicle arrives (in hours). Find the probability the station will have to wait at least 30 minutes for the first vehicle to arrive. Note that 30 minutes is 1/2 of an hour.
(c) (Chapter 4) What distribution describes the time until the 4th vehicle arrives?

3.11. Let X denote the number of requests for assistance received by a towing service per hour. Suppose X follows a Poisson distribution with mean $\lambda = 5$.
(a) Hourly company revenue R (in dollars) has been modeled as a quadratic function of X; specifically, $R = -120 + 40X + 10X^2$. Calculate the expected hourly revenue $E(R)$.
(b) (Chapter 4) Let W denote the time it takes for the service to receive its first call. Name the distribution of W and find the probability the service will have to wait longer than 10 minutes to receive the first call. Note that 10 minutes is 1/6th of an hour.

CHAPTER 4:

4.1. The amount of gravel (in tons) sold by a construction company on a given day is modeled as a continuous random variable X with probability density function (pdf):

$$f_X(x) = \begin{cases} 0.02(10 - x), & 0 < x < 10 \\ 0, & \text{otherwise.} \end{cases}$$

(a) What is the probability the company will sell less than 5 tons on a given day?
(b) Calculate $E(X)$ and give an interpretation of what it means.
(c) Find the median amount of gravel sold on a given day. Interpret what this means.

4.2. A mechanic records X, the amount of time during a one-hour period a machine operates at its maximum capacity. Possible values of X are between 0 and 1; that is, "0" means the machine never operates at maximum capacity during the hour and "1" means the machine operates at full capacity during the entire hour. The random variable X is continuous and has the following probability density function (pdf):

$$f_X(x) = \begin{cases} 4x(1 - x^2), & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Calculate $E(X)$ and $V(X)$.
(b) Let $F_X(x)$ denote the cumulative distribution function (cdf) of X. Calculate $F_X(0.5)$.

4.3. "Time headway" in traffic flow is the elapsed time between when one car completely passes a fixed point and when the next car begins to pass the same point. Let X denote this elapsed time (in seconds) for traffic on I-77 in Rock Hill, SC, during peak driving times. Engineers model X using the probability density function (pdf):

$$f_X(x) = \begin{cases} \dfrac{2}{x^3}, & x > 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Calculate the probability the time headway X will be less than 3 seconds.
(b) Calculate $E(X)$, the mean time headway.
(c) Find the cumulative distribution function $F_X(x)$ and graph it. On the horizontal axis, use a graphing range of $(0, 6)$ with tick marks at $0, 1, 2, ..., 6$.

4.4. Explosive devices used in mining operations produce circular craters when detonated. The radii of these craters X follow an exponential distribution with $\lambda = 0.2$ meters.
(a) Find the probability that a single crater's radius will be between 5 and 15 meters.
(b) Sketch a graph of the probability density function (pdf) of X and indicate on the graph the probability you calculated in part (a). Label axes. Neatness counts.
(c) The area of a crater with radius X is $W = \pi X^2$. Calculate the expected area $E(W)$.

4.5. A gas station is supplied with gasoline once per day. The station's daily sales (in 10,000s of gallons) is a continuous random variable X with probability density function (pdf)

$$f_X(x) = \begin{cases} 0.25(2-x)^3, & 0 < x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find $P(X < 0.4)$, the probability daily sales are less than 4000 gallons.
(b) The station owner wants to determine the supply c so that the probability the daily sales X exceeds c is 0.01. Determine what c is. *Hint:* What quantile (percentile) of X does the owner want to determine?

4.6. During an 8-hour shift, the proportion of time X a sheet-metal stamping machine is operating is a continuous random variable with probability density function (pdf):

$$f_X(x) = \begin{cases} 6x^5, & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the probability the machine is operating for at least 90% of the shift, that is, calculate $P(X > 0.9)$.
(b) Calculate $E(X)$, the expected proportion of time the machine is operating during an 8-hour shift.
(c) Calculate the variance of X.

CHAPTER 5:

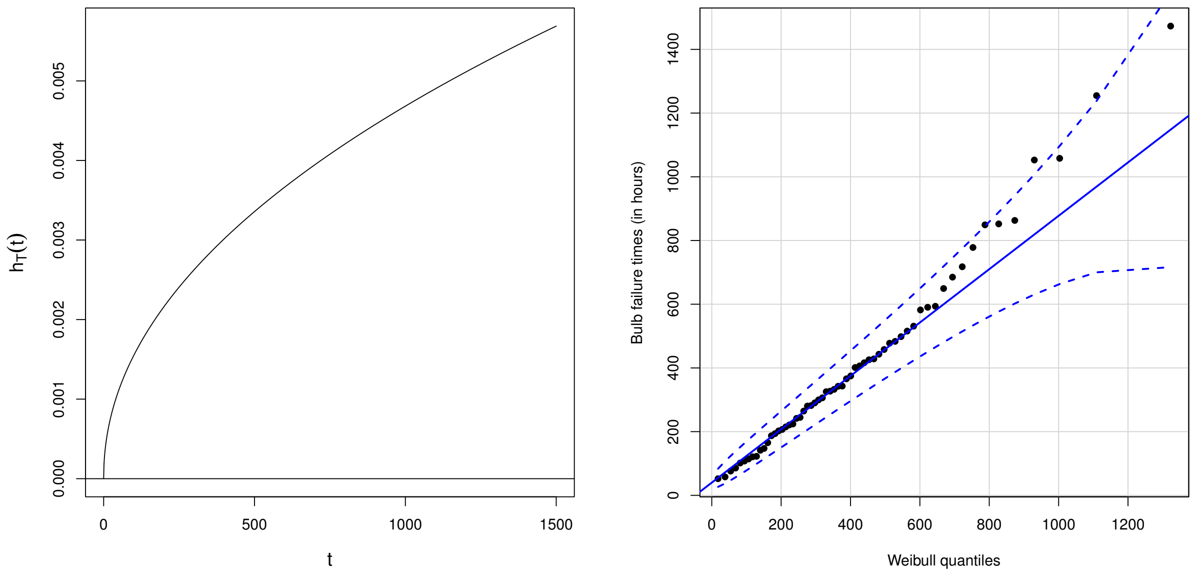
5.1. Resistors used in the construction of an aircraft guidance system have lifetimes T (in 100s of hours) that are modeled using a Weibull distribution. From years of historical data, engineers use $\beta = 1.5$ and $\eta = 150$.
(a) Calculate the probability a randomly selected resistor will have a lifetime between 10,000 and 20,000 hours. That is, calculate $P(100 < T < 200)$.
(b) Ninety percent of resistors will fail before what time?
(c) Recall that the hazard function of T is

$$h_T(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}.$$

Graph the hazard function in this example and interpret what it means.

5.2. A light bulb company manufactures filaments that are not expected to wear out during an extended period of "intense use." With the goal of guaranteeing bulb reliability in these conditions, engineers sample $n = 60$ bulbs, simulate their long-term intense use, and record T, the hours until failure for each bulb. Here are the data:

443.0	593.5	374.9	582.0	590.4	290.4	264.6	649.3	531.1	849.2
101.5	107.7	141.5	342.4	122.5	401.3	57.9	147.2	281.9	852.2
52.3	477.6	85.7	221.1	685.0	343.1	187.1	515.7	202.3	1058.0
498.2	241.6	244.7	1052.7	406.4	165.2	193.7	425.7	76.2	416.2
457.9	778.1	483.4	224.2	325.4	1254.9	280.3	206.8	717.6	863.0
327.0	332.7	214.9	121.0	428.3	306.2	1473.1	365.9	114.3	299.7



Problem 5.2. Left: Estimated hazard function. Right: Weibull qq plot.

The engineers assume a Weibull(β, η) model for T. Here are the maximum likelihood estimates of β and η from fitting the model to the data:

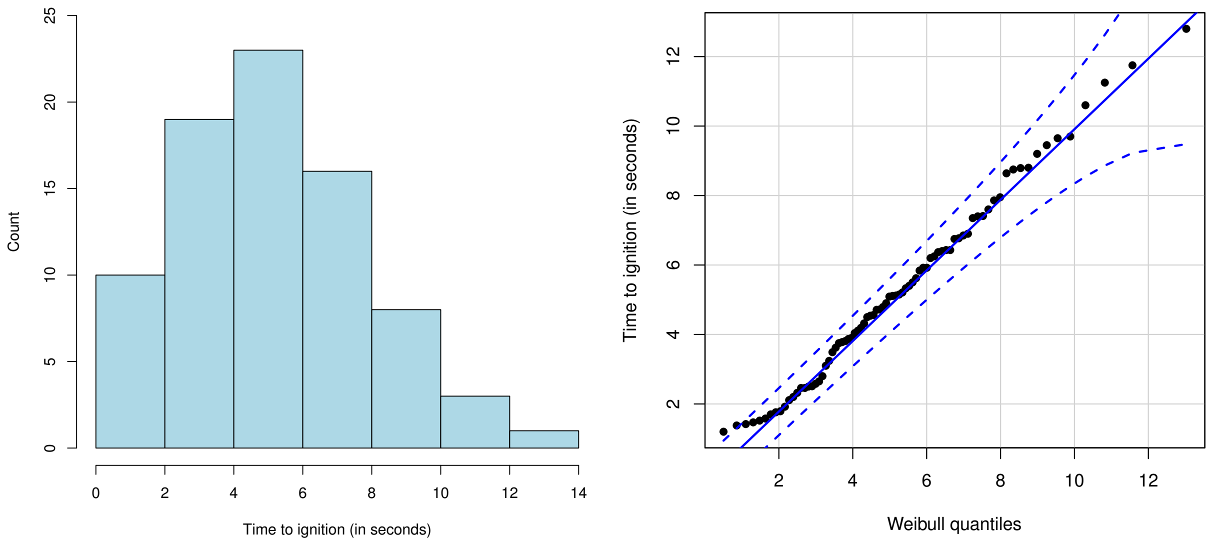
$$\widehat{\beta} = 1.5 \qquad \widehat{\eta} = 459.2.$$

(a) Calculate an estimate of $\phi_{0.10}$, the 10th percentile of Weibull population distribution.
(b) The estimate of β is larger than 1, which means the estimated hazard function $h_T(t)$ is an increasing function of t. I have plotted this function above (left). Interpret what it means.
(c) The quantile-quantile (qq) plot from the Weibull model fit is also shown above (right). Interpret this plot. Should the engineers be concerned about using the Weibull distribution to assess the reliability of these bulbs?

5.3. In a fire safety study, engineers collected a random sample of $n = 80$ upholstery pieces and on each piece recorded T, the ignition time (in seconds) when exposed to a flame. Here are the data:

2.58	2.51	4.04	6.43	1.58	4.32	2.20	4.19	4.79	6.20
1.52	1.38	3.87	4.54	5.12	5.15	5.50	5.92	4.56	2.46
6.90	1.47	2.11	2.32	6.75	5.84	8.80	7.40	4.72	3.62
2.46	8.75	2.65	7.86	4.71	6.25	9.45	12.80	1.42	1.92
7.60	8.79	5.92	9.65	5.09	4.11	6.37	5.40	11.25	3.90
5.33	8.64	7.41	7.95	10.60	3.81	3.78	3.75	3.10	6.43
1.70	6.40	3.24	1.79	4.90	3.49	6.77	5.62	9.70	5.11
4.50	2.50	5.21	1.76	9.20	1.20	6.85	2.80	7.35	11.75

A histogram of the data and a quantile-quantile plot under a Weibull assumption are shown on the top of the next page.



Problem 5.3. Left: Histogram. Right: Weibull qq plot.

(a) Do you think the Weibull model is a reasonable choice for the population-level distribution of ignition time? Explain.
(b) When I fit a Weibull(β, η) distribution to these data, R reported the maximum likelihood estimates

$$\begin{aligned}\widehat{\beta} &= 2.05 \\ \widehat{\eta} &= 5.91.\end{aligned}$$

Use these values to estimate the proportion of upholstery pieces in the population that will ignite before 5 seconds. *Hint:* I'm asking you to find $P(T < 5)$ using $\widehat{\beta}$ and $\widehat{\eta}$ as estimates of β and η, respectively.
(c) What does the survivor function $S_T(t)$ represent in this example?

5.4. A researcher wants to model the time until failure for a new cooling unit specifically designed for wine cellars. He has access to the manufacturer's published data on failure times for $n = 25$ units tested; these data are below:

0.32	1.15	1.43	1.47	1.60	1.62	2.12	2.38	2.50	2.82	2.82	2.98	3.00
3.02	3.19	3.44	3.77	3.79	3.89	3.99	4.07	4.10	4.17	4.18	4.19	

These data are measurements of T, the time (in years) until the cooling unit fails. He assumes a Weibull(β, η) distribution for T. Here are the maximum likelihood estimates of the shape (β) and scale (η) parameters in the Weibull distribution based on the data above:

```
> fitdistr(failure.times,densfun="weibull")
  shape     scale
   2.92      3.22
  (0.50)    (0.23)
```

(a) With the maximum likelihood estimates $\widehat{\beta} \approx 2.92$ and $\widehat{\eta} = 3.22$,

- estimate the percentage of cooling units in the population that will be operational at 5 years.
- estimate $\phi_{0.5}$, the median time until failure.

(b) (Chapter 6) The quantities in parentheses above (0.50) and (0.23) are the standard errors of the maximum likelihood estimates. Explain to the researcher what these measure.
(c) Suppose the Weibull distribution is an excellent model for the failure times of the cooling units. Sketch a graph of what the qq plot might look like for the data above. Label both axes and choose suitable scales for each axis. Neatness counts.

5.5. Time to event studies are common in medical applications. In many of these studies, the event of interest means "death" from a serious disease. However, in other studies, the event is something positive. Consider a recent study involving patients with venous ulcers (also known as leg ulcers). For one group of $n = 187$ patients, a short-stretch bandage was applied to each patient's infected leg area. The time to event measured on each patient was

$$T = \text{time (in days) until the leg ulcer was completely healed.}$$

Under a Weibull model assumption for T, I estimated the parameters β and η using maximum likelihood; here is the R output:

```
> fitdist(healing.times,"weibull")
Parameters:
      estimate Std.Error
shape   0.999     0.056
scale 190.987    14.794
```

The estimates of β (shape) and η (scale) are $\widehat{\beta} \approx 1$ and $\widehat{\eta} \approx 191$, respectively.
(a) What distribution is a special case of the Weibull when the shape parameter $\beta = 1$? If β really was 1, what would this imply about healing rate in this population of patients over time? Explain.
(b) Under the estimated Weibull model with $\beta = 1$ and $\eta = 191$, calculate the median healing time $\phi_{0.5}$. Interpret what this means.
(c) What does the quantile-quantile plot (above) suggest about the Weibull distribution as a model for these data?
(d) Name another lifetime distribution that might be used to model the healing times in this example.

5.6. Mechanical engineers are studying the reliability of industrial strength air conditioning units. They have access to a random sample of $n = 42$ units and observe

$$T = \text{time until failure (in 1000s of hours)}$$

for each unit. Under a Weibull(β, η) assumption for the population distribution of T, the maximum likelihood estimates of β and η based on the sample are shown on the next page:

```
> fitdist(failure.time,distr="weibull",method="mle")

      estimate  Std.Error
shape     1.04      0.126
scale    22.80      3.558
```

The maximum likelihood estimates are

$$\widehat{\beta} = 1.04$$
$$\widehat{\eta} = 22.80.$$

The standard error of each estimate is also shown in the output above.
(a) Using the estimates of β and η above, find the probability a unit fails between 40,000 and 60,000 hours. That is, find $P(40 < T < 60)$.
(b) (Chapter 7) Mathematical arguments show that maximum likelihood estimates are approximately normally distributed in large samples. This means the formula

$$\texttt{estimate} \pm 1.96(\texttt{Std.Error})$$

can be used to write a 95% confidence interval. Using the output above, calculate this interval for the population shape parameter β and interpret it.
(c) (Chapter 7) One of the engineers says,

> *"I knew β would be larger than 1. This means the population of units is getting weaker over time."*

Do you agree with this statement? Explain.

PMF AND PDF FORMULAS

Binomial:

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, ..., n \\ 0, & \text{otherwise.} \end{cases}$$

Geometric:

$$p_X(x) = \begin{cases} (1-p)^{x-1} p, & x = 1, 2, 3, ... \\ 0, & \text{otherwise.} \end{cases}$$

Negative binomial:

$$p_X(x) = \begin{cases} \binom{x-1}{r-1} (1-p)^{x-r} p^r, & x = r, r+1, r+2, ..., \\ 0, & \text{otherwise.} \end{cases}$$

Hypergeometric:

$$p_X(x) = \begin{cases} \dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, & x \le K \text{ and } n - x \le N - K \\ 0, & \text{otherwise.} \end{cases}$$

Poisson:

$$p_X(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, ... \\ 0, & \text{otherwise.} \end{cases}$$

Exponential:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise.} \end{cases} \qquad F_X(x) = \begin{cases} 0, & x \le 0 \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

Gamma:

$$f_X(x) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Normal (Gaussian):

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \, e^{-(x-\mu)^2/2\sigma^2}, \quad \text{for } -\infty < x < \infty.$$

Weibull:

$$f_T(t) = \begin{cases} \dfrac{\beta}{\eta}\left(\dfrac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\dfrac{t}{\eta}\right)^{\beta}\right], & t > 0 \\ 0, & \text{otherwise.} \end{cases} \qquad F_T(t) = \begin{cases} 0, & t \le 0 \\ 1 - \exp\left[-\left(\dfrac{t}{\eta}\right)^{\beta}\right], & t > 0. \end{cases}$$