1. 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. 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. 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.

4. 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) 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) What distribution describes the time until the 4th vehicle arrives? Give me the name of the distribution and the values of the parameters in it.

5. 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.

6. 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.

7. 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?

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?

9. 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) 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.

10. 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 $F_X(0.5)$.

11. 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

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 on the next page (left). Interpret what it means.
(c) The quantile-quantile (qq) plot from the Weibull model fit is also shown on the next page (right). Interpret this plot. Should the engineers be concerned about using the Weibull distribution to assess the reliability of these bulbs?

12. 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).

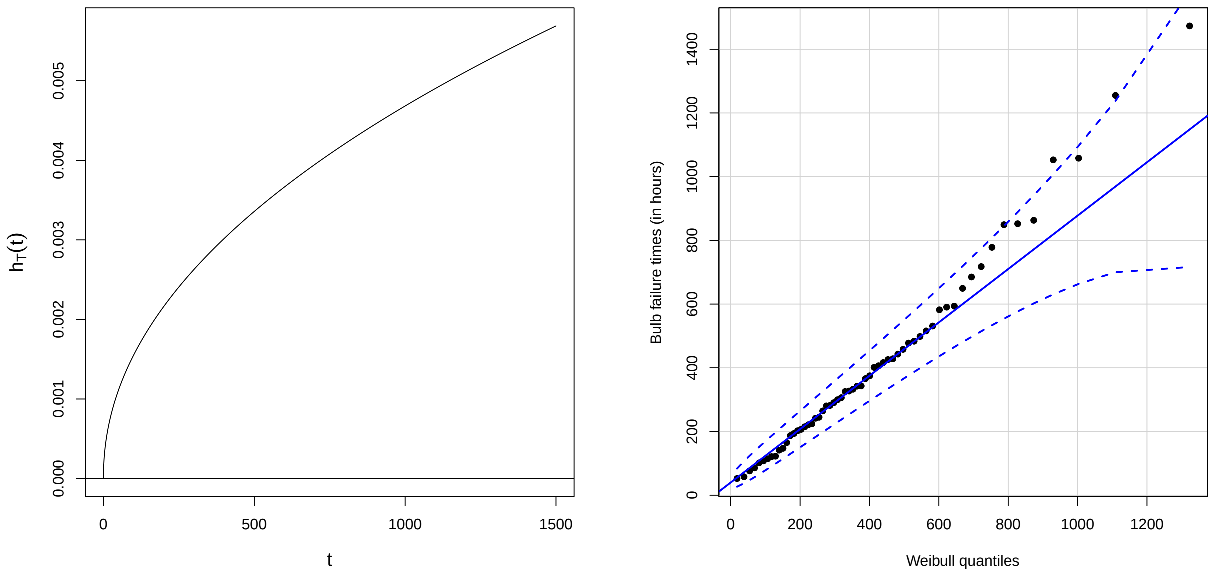

Figure 1: Problem 11. Left: Estimated hazard function. Right: Weibull qq plot.

(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?

13. "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 $0, 1, 2, ..., 6$.

14. 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

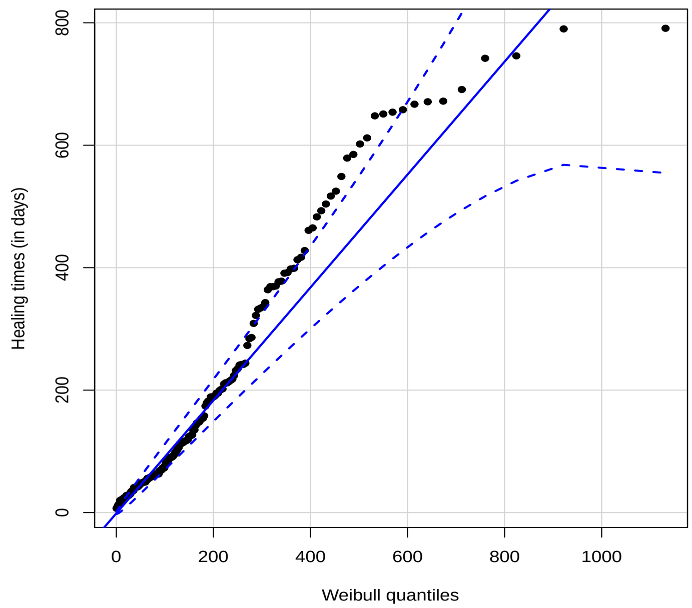

Figure 2: Problem 14. Quantile-quantile plot for healing time data.

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.

15. 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 that all batteries are independent.

16. 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 $E(W)$.

17. A viticulturist 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) The quantities in parentheses above (0.50) and (0.23) are the standard errors of the maximum likelihood estimates. Explain to the viticulturist 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.

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:

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

Poisson:

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

Exponential:

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

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:

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