STAT 687 Homework #1 - Solution

From the text problem # 1.1

Discuss the assumptions that would be needed to convert the heat exchanger data from calendar time in Figure 1.6 to operating-time data in Figure 1.7 and then use these data for purposes of analysis and inference on the life distribution of heat exchanger tubes of the type in these heat exchangers.

The phrase ". . . analysis and inference on the life distribution. . . " implies a common life distribution is desired to make statements on the lifetime of the overall population of heat exchanger tubes based on the combined data obtained at each of the three plants.

Assumptions:

- Though the tubes were probably manufactured at different times, we would have to assume that the tubes installed in the three plants were produced using manufacturing processes that could be considered approximately the same.
- No external events (e.g., earthquake, maintenance actions) that could cause failures occurred at a particular plant during the observation period.
- All three plants were run under similar environmental conditions and similar workloads for the entire period of observation.

From the text problem # 1.3

In the development and presentation of traditional statistical methods, description and inference are often represented in terms of means and variances (or standard deviations) of distributions.

(a) *Use some examples in this chapter to explain why in many applications reliability or design engineers would be more interested in the time at which 1% (or some smaller percentage) of a particular component will fail instead of the time at which 50% would fail.*

Nearly all products will be taken out of service long before the mean of the failure time distribution is reached. For the heat exchanger tube and turbine wheel examples, safety is a major concern. Therefore, reaching the number of failures that represents the mean of the failure time distribution is unacceptable. Rather, the focus is on ensuring that the probability of a single failure is low over a reasonable design life or until a scheduled inspection occurs.

For the Ni-Cd battery and printed circuit board examples minizing the number of warranty claims may be the driving factor. For these products a 5-10% return rate during the warranty period could mean a significant loss of both profit and market share to a competitor.

(b) *Explain why means and variances of time to failure may not be of such high interest in reliability studies*

In reliability practice, the failure time distribution may be highly skewed. For such distributions, the mean and variance are less informative. Further, many important reliability questions are often answered in terms of probabilities (e.g., the probability of survival or failure by a certain time or after a certain amount of use) the hazard function (the probability of failure in the next small interval of time, given survival to that time), and quantiles (the times at which a specified proportion of a large population will have failed).

(c) *Give at least one example of a product for which the mean time to failure would be of interest. Explain why.*

The mean time to failure provides an economically sensible reliability metric for products that may fail during service with little or no adverse consequences and when all or most units are expected to fail in service (consumable items). An example would be light bulbs or automobile tires. For such products the cost of operation (including replacements) is inversely proportional to the mean time to failure.

From the text problem # 1.5

An important part of quantifying product reliability is the specification of an appropriate time-scale (or time-scales) on which life should be measured (e.g. hours of operation, cycles of operation, etc.). For each of the following products, suggest and give reasons for an appropriate scale (or scales) upon which one might measure life. Also, discus possible environmental factors that might affect the lifetime of individual units.

It is important to consider the failure mechanism when determining an appropriate time scale. The time scale should have a direct relationship to the way the unit under consideration progresses toward failure. Some products do not have a simple single time scale (often because there is more than one underlying failure mechanism). See the answers to 1.6.

From the text problem # 1.6

For each of the products listed in Exercise 1.5 explain your best understanding of the underlying failure mechanism. Also, describe possible ways in which an analyst could define failure.

(a) **Painted surface of an automobile** – Today's automobile coatings are complicated systems consisting of a base primer coat, paints with color, and a clearcoat top layer. These coating systems degrade chemically, especially when exposed to UV (ultraviolet) radiation. The degradation is accelerated with higher temperatures, humidity, and cycling of these variables. UV degradation can be retarded by putting UV absorbers in the coating mixtures. Defects in application, such as thin spots, can cause especially early failures. In controlled testing, the amount of UV exposure would be the most accurate time scale. When gathering customer data, age of the automobile, part of the country where it is typically driven, and parking habits of the owner/user might provide reasonable surrogate information. Failure in this case might be defined by the diameter of a chip in the paint or the number of paint chips over a defined surface area. Finally, failure may be defined as the point when the gloss on the finish reaches a certain level of dullness.

(b) **Automobile lead-acid battery** – Standard automobile lead-acid batteries for automobiles powered by internal-combustion engines are designed to serve as float batteries and also have a large amount of power to start the engine. Most of the power to actually run the automobile comes from the alternator. Although lead-acid batteries will degrade over time even if not discharged, these batteries suffer especially heavy degradation when they endure deep discharge-charge cycles. Cold weather is more stressful for such batteries because less charge can be held in lower temperatures, more power is needed to turn an engine that is cold, and engines are harder to start when it is cold. Batteries that are designed to provide the main source of power for vehicles (lead-acid and other types) are designed to withstand larger numbers of charge/discharge cycles, but would have less power available for starting a large engine. Charge-discharge cycles are still the appropriate time scale. Again, in controlled testing, the number of charge-discharge cycles would be an appropriate scale in which to measure degradation and life of a rechargeable battery. Failure of a battery is typically defined as the time at which the voltage in the weakest cell falls below a certain threshold.

(c) **Automobile windshield wipers** – The rubber in automobile windshield wipers degrades over time as result of exposure to an external environment. This degradation can be accelerated by UV exposure. The proper time scale in this case would be the number of use-cycles (passes across the windshield), By establishing a baseline environment (lab environment, for example) UV exposure cn be included as an accelerating factor depending on the amount of use relative to real time and the amount of UV

exposure. A common practice with such products is to "age" them at high levels of temperature and UV and then to test life or durability of the aged units.

(d) **Automobile tires** – For the most part, tires wear as a function of the number of miles driven. The wear rate may be affected by temperature, driving style, age of the tire, over or under inflation, etc. Tires also degrade chemically over time, especially when exposed to UV. Failure of a tire is usually defined as the point at which the tred has worn to a specified level. Tires become more brittle, due to ageing chemical degradation. In the process, they sometimes develop cracks, especially in the sidewall. Of course tires can also fail from puncture or other accidents.

(e) **Incandescent light bulb** – The filaments in an operating incandescent light bulb will evaporate over time. When a bulb is turned on and off, thermal gradients cause mechanical shocks that can lead to fatigue. It is usually such a shock that causes a light bulb to fail (you may have noticed that light bulbs are more likely to fail when you first turn them on). Vibration can also accelerate fatigue failure. Failure is usually simple to define in this case: the filament burns out and bulb stops functioning.

From the text problem # 1.8

For each of the products listed in Exercise 1.7, describe the range of environments that the product might encounter in use and the effect that environment could have on the product's reliability

(a) **Automobile alternator** – Elevated temperatures during operation, low temperatures at start up during the Winter months, signigicant vibration loads from engine operation and impact with potholes, speedbumps, etc., hundreds of thermal cycles between 0^oF at start up and 200^oF during operation, exposure to corrosive environments from spilled/leaking engine fluids as well as spray from road moisture combined with snow melt chemicals.

(b) **Video cassette recorder** – Drops during shipment, high use patterns, electrical surges, plugging into receptacle with improper home wiring, improper usage (inserting foreign objects)

(c) **Microwave oven** – Drops during shipment, high use patterns, electrical surges, plugging into receptacle with improper home wiring, improper use (putting things in that don't belong)

(d) **Home air conditioner** – High usage during Summer months, dormant failures from cold exposure during Winter, impact from objects (toys, balls, tree limbs)

(e) **Hand-held calculator** – Impacts from dropping on floor or into a bookbag, high usage

(f) **Clothes dryer** – High usage, improper maintenace (not cleaning lint trap), excessive usage (routinely putting larger loads than designed), improper usage (inserting foreign objects)

From the text problem # 1.14

There was a considerable amount of censoring at low levels of humidity in the printed circuit accelerated life test data shown in Figure 1.9. Explain how such censoring can obscure important information about the relationship between humidity and time to failure.

In an accelerated life test our interest is in translating the observed times to failure at one or more elevated stress levels to some use-level stress. Usually, this is accomplished by extrapolating a regression line through a specific life quantile (mean life, median life, characteristic life, etc.) at each elevated stress level to that same life quantile at the use stress.

If a signifigant number of observations obtained at an elevated stress are censored, the estimate of the mean, median or characteristic life may have a high level of uncertainty. As result, there may be a high level of uncertainty on the slope of regression line, leading to significant errors in the life prediction at the use level stress. For this example, a majority of the systems tested at the lowest humidity level did not fail prior to the end of the test. The time t_p corresponding to the $100p\%$ fraction of failures at this humidity level would be overly conservative.

From the text problem # 2.2

It is possible for a continuous cdf to be constant over some intervals of time

(a) *Give an example of a physical situation that would result in a cdf, $F(t)$ that is constant over some values of t.*

If life time is measured in real time, and if units do not operate during some times of the life-cycle, then there will be flat spots in the cdf during the off-times. For example, a light bulb that is turned off 12 hours each day will only fail during the operating hours. Yet lifetime may be measured in terms of days since the bulb was installed. If flat spots exist in the cdf of a product's lifetime distribution, it is likely an indication that the definition of time scale could be improved. For the light bulb example, using operating time as the lifetime measure would be more appropriate than calendar time.

(b) *Sketch such a cdf and its corresponding pdf*

Suppose that the lifetime of the lightbulbs discussed in part (a) follows a *Lognor*$(5, .85)$ distribution (this is a poor assumption since a lightbulb's liftime is one of the few things that is known to follow an exponential distribution). Under this assumption, if the lightbulbs were used continuously (24-hrs/day) the respective pdf and cdf plots would appear as shown in Figure 1.

```
points<-seq(0,480)
plot(points,dlnorm(points,5,.85), type="l", lwd = 2, yaxt = "n",
     xlab = "Hours", ylab = "f(t)")
box(lwd = 1.5)
axis(side = 2, las = 1, tck = 0.04, hadj=0.65)

plot(points,plnorm(points,5,.85), type="l", lwd = 2, yaxt = "n",
     xlab = "Hours", ylab = "f(t)")
box(lwd = 1.5)
axis(side = 2, las = 1, tck = 0.04, hadj=0.65)
```

If usage occurs in 12-hour increments and the reliability metric is stated with respect to hours, the cdf and pdf plots would appear as shown in Figure 2.

```
points<-seq(0,480,12)
p<-points[c(1,rep(2:70,each = 2),71)]
```
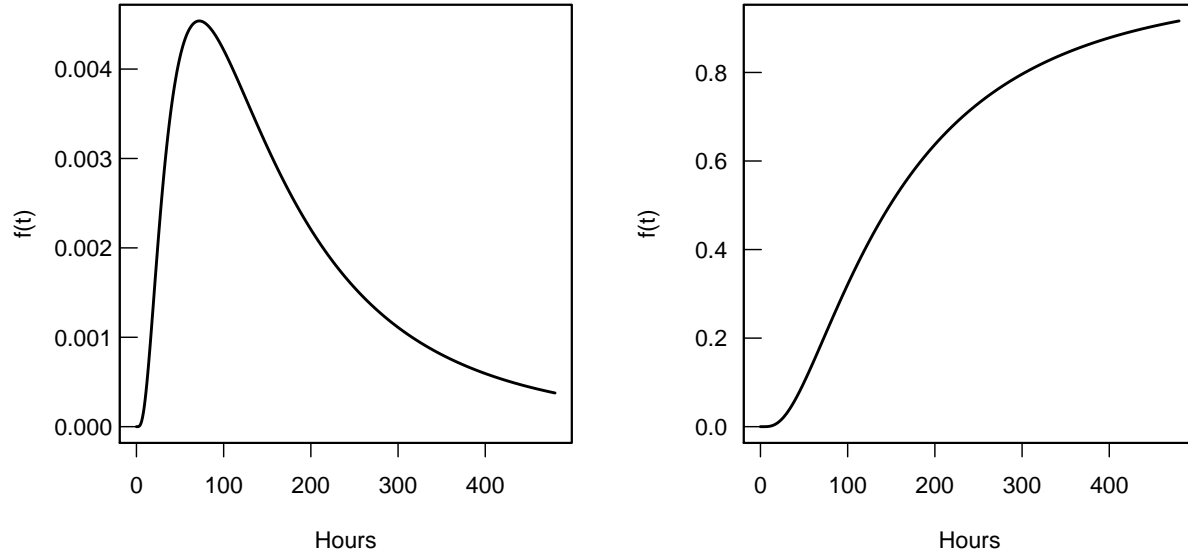
Figure 1: Pdf (left) and cdf curves for a lognormal(5,.85) distribution assuming continuous (24-hr) usage

```
Z<-function(a) rep(plnorm(p[a],5,.85),13)
z<-rep(0,13)
par(mfrow = c(1,2))
plot(
c(p[ 1]:p[ 2],p[3] :p[ 4],p[ 5]:p[ 6],p[ 7]:p[ 8],p[ 9]:p[10],p[11]:p[12],p[13]:p[14],
  p[15]:p[16],p[17]:p[18],p[19]:p[20],p[21]:p[22],p[23]:p[24],p[25]:p[26],p[27]:p[28],
  p[29]:p[30],p[31]:p[32],p[33]:p[34],p[35]:p[36],p[37]:p[38],p[39]:p[40],p[41]:p[42],
  p[43]:p[44],p[45]:p[46],p[47]:p[48],p[49]:p[50],p[51]:p[52],p[53]:p[54],p[55]:p[56],
  p[57]:p[58],p[59]:p[60],p[61]:p[62],p[63]:p[64],p[65]:p[66],p[67]:p[68],p[69]:p[70]),

c(dlnorm(p[ 1]:p[ 2],5,.85),z,dlnorm(p[ 3]:p[ 4],5,.85),z,dlnorm(p[ 5]:p[ 6],5,.85),z,
  dlnorm(p[ 7]:p[ 8],5,.85),z,dlnorm(p[ 9]:p[10],5,.85),z,dlnorm(p[11]:p[12],5,.85),z,
  dlnorm(p[13]:p[14],5,.85),z,dlnorm(p[15]:p[16],5,.85),z,dlnorm(p[17]:p[18],5,.85),z,
  dlnorm(p[19]:p[20],5,.85),z,dlnorm(p[21]:p[22],5,.85),z,dlnorm(p[23]:p[24],5,.85),z,
  dlnorm(p[25]:p[26],5,.85),z,dlnorm(p[27]:p[28],5,.85),z,dlnorm(p[29]:p[30],5,.85),z,
  dlnorm(p[31]:p[32],5,.85),z,dlnorm(p[33]:p[34],5,.85),z,dlnorm(p[35]:p[36],5,.85)),
    type="l", lwd = 2, yaxt = "n", xlab = "Hours", ylab = "f(t)")
box(lwd = 1.5)
axis(side = 2, las = 1, tck = 0.04, hadj=0.65)




plot(
c(p[ 1]:p[ 2],p[3] :p[ 4],p[ 5]:p[ 6],p[ 7]:p[ 8],p[ 9]:p[10],p[11]:p[12],p[13]:p[14],
  p[15]:p[16],p[17]:p[18],p[19]:p[20],p[21]:p[22],p[23]:p[24],p[25]:p[26],p[27]:p[28],
  p[29]:p[30],p[31]:p[32],p[33]:p[34],p[35]:p[36],p[37]:p[38],p[39]:p[40],p[41]:p[42],
  p[43]:p[44],p[45]:p[46],p[47]:p[48],p[49]:p[50],p[51]:p[52],p[53]:p[54],p[55]:p[56],
```

```
   p[57]:p[58],p[59]:p[60],p[61]:p[62],p[63]:p[64],p[65]:p[66],p[67]:p[68],p[69]:p[70]),

c(plnorm(p[ 1]:p[ 2],5,.85),Z( 2),plnorm(p[ 3]:p[ 4],5,.85),Z( 4),
  plnorm(p[ 5]:p[ 6],5,.85),Z( 6),plnorm(p[ 7]:p[ 8],5,.85),Z( 8),
  plnorm(p[ 9]:p[10],5,.85),Z(10),plnorm(p[11]:p[12],5,.85),Z(12),
  plnorm(p[13]:p[14],5,.85),Z(14),plnorm(p[15]:p[16],5,.85),Z(16),
  plnorm(p[17]:p[18],5,.85),Z(18),plnorm(p[19]:p[20],5,.85),Z(20),
  plnorm(p[21]:p[22],5,.85),Z(22),plnorm(p[23]:p[24],5,.85),Z(24),
  plnorm(p[25]:p[26],5,.85),Z(26),plnorm(p[27]:p[28],5,.85),Z(28),
  plnorm(p[29]:p[30],5,.85),Z(30),plnorm(p[31]:p[32],5,.85),Z(32),
  plnorm(p[33]:p[34],5,.85),Z(34),plnorm(p[35]:p[36],5,.85)),

type="l", lwd = 2, yaxt = "n", xlab = "Hours", ylab = "F(t)")
box(lwd = 1.5)
axis(side = 2, las = 1, tck = 0.04, hadj=0.65)
```
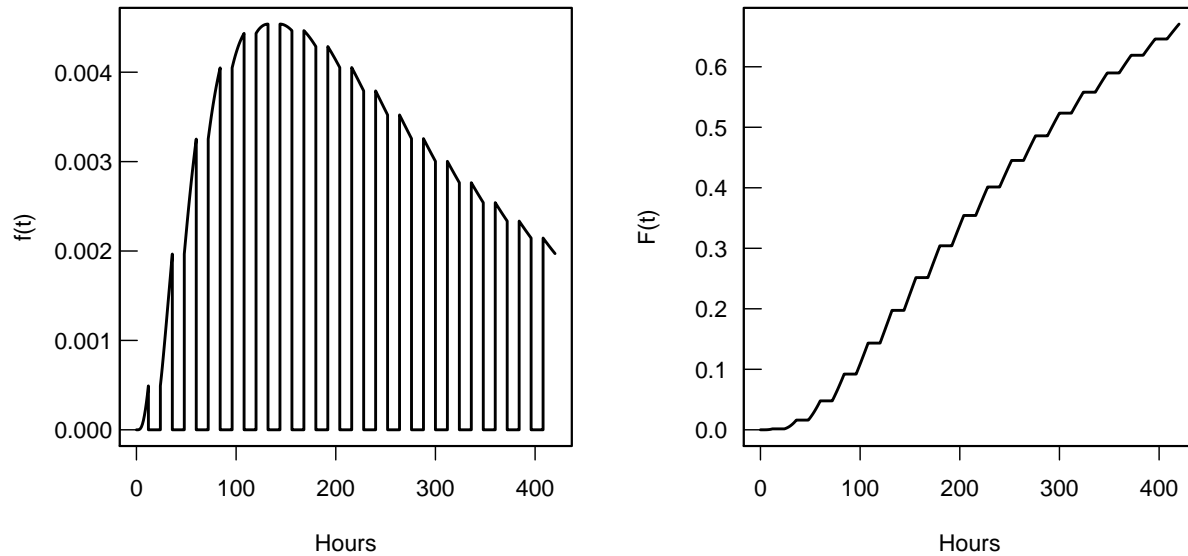

Figure 2: Pdf (left) and cdf curves for a lognormal(5,.85) distribution assuming half-day usage (12-hr intervals)

(d) *Think about and explain the relationship between the occurrence of flat spots in a cdf and the choice of an approriate time scale on which to define life.*

Continuing with the lightbulb example, the flat spots in the cdf shown in Figure 2 occur because the time unit used to generate the plot differs from the time unit used to measure usage. Had "operating hours" been used instead of "hours", the cdf plot would appear the same as shown in Figure 1.

6

From the text problem # 2.4

Consider a cdf $F(t) = 1 - \exp\left[-\frac{t}{\eta}\right]^{\beta}, t > 0, \eta > 0, \beta > 0$. (This is the cdf of the Weibull distribution, which will be discussed in detail in Chapter 4.)

   (a) *Derive an expression for the pdf $f(t)$.*

$$f(t) = \frac{d}{dt}F(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\}$$

   (b) *Derive an expression for the hazard function $h(t)$.*

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$$

   (c) *Sketch (or use a computer to draw) the cdf, pdf, and hazard functions for $\eta = 1$ and $\beta = .5, 1, 2$.*

The plots are shown in Figure 3 below and may be produced using the following R code:

```
par(mfrow=c(3,3), family="serif")
weihaz <- function(x,a,b) {(a/b)*(x/b)^(a-1)}
curve(dweibull(x,.5,1),xlim=c(0,6),ylab="",xlab="PDF of Weibull(.5,1)")
curve(dweibull(x,1,1),xlim=c(0,6),ylab="",xlab="PDF of Weibull(1,1)")
curve(dweibull(x,2,1),xlim=c(0,6),ylab="",xlab="PDF of Weibull(2,1)")
curve(pweibull(x,.5,1),xlim=c(0,6),ylim=c(0,1),ylab="",xlab="CDF of Weibull(.5,1)")
curve(pweibull(x,1,1),xlim=c(0,6),ylim=c(0,1),ylab="",xlab="CDF of Weibull(1,1)")
curve(pweibull(x,2,1),xlim=c(0,6),ylim=c(0,1),ylab="",xlab="CDF of Weibull(2,1)")
curve(weihaz(x,.5,1),xlim=c(0,6),ylim=c(0,2),ylab="",xlab="Hazard of Weibull(.5,1)")
curve(weihaz(x,1,1),xlim=c(0,6),ylim=c(0,2),ylab="",xlab="Hazard of Weibull(1,1)")
curve(weihaz(x,2,1),xlim=c(0,6),ylab="",xlab="Hazard of Weibull(2,1)")
```
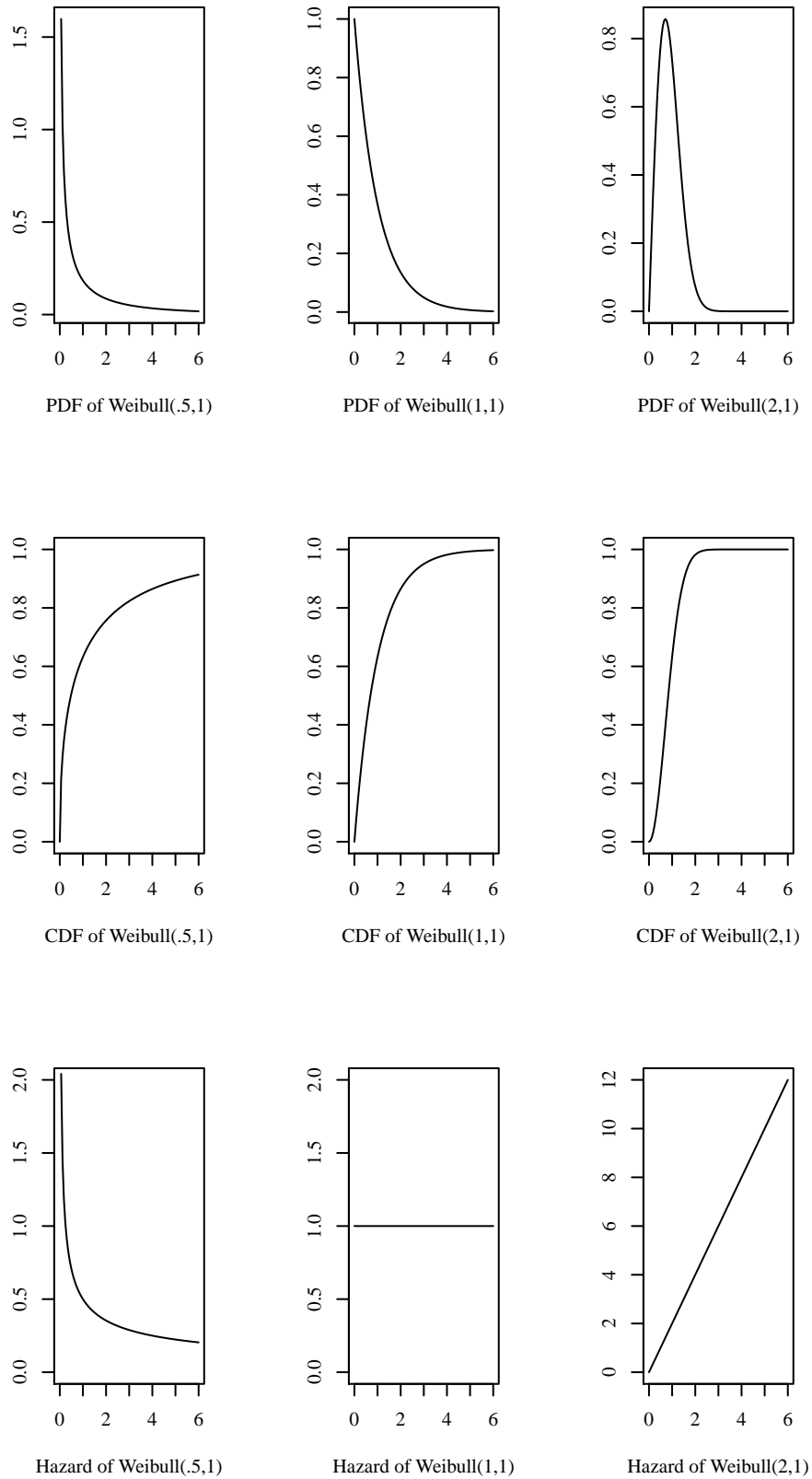
PDF of Weibull(.5,1)

PDF of Weibull(1,1)

PDF of Weibull(2,1)

CDF of Weibull(.5,1)

CDF of Weibull(1,1)

CDF of Weibull(2,1)

Hazard of Weibull(.5,1)

Hazard of Weibull(1,1)

Hazard of Weibull(2,1)

Figure 3: Pdf, CDF and hazard function curves for various Weibull distributions

From the text problem # 2.13

Show that the pdf, cdf, survival, hazard, and cumulative hazard are mathematically equivalent descriptions of a continuous distribution in the sense that given any of these functions the other four are completely determined

Need to show that if we start with any one of these functions we can obtain the others from it, i.e.

$$f_X(x) \rightarrow F_X(x) \rightarrow S_X(x) \rightarrow h_X(x) \rightarrow H_X(x) \rightarrow f_X(x)$$

Let's begin with the pdf, $f_X(x)$.

$$F_X(x) = \int_0^x f_X(t)\, dt$$

$$S_X(x) = 1 - F_X(x)$$

$$h_X(x) = \frac{d}{dx} log\left(S_X(x)\right)$$

$$H_X(x) = \int_0^x h_X(t)\, dt$$

$$f_X(x) = -\frac{d}{dx} \exp\left\{-H_X(x)\right\}$$

Since we can obtain any of these functions regardless of which function we are initially given, we have obtained the desired result.

From the text problem # 2.16

If a continuous random variable T has a cdf $F(t) = Pr(T \leq t)$, then it is easy to show that the transformed random variable $F(T)$ follows a $UNIF(0,1)$ distribution. A similar property for random variables is that the cumulative hazard transformation $H(T)$ follows an exponential distribution. Show this.

The first sentence in this problem states that given a random variable T with continuous distribution function $F_T(t)$, the transformed random variable $U = F_T(T) \sim U(0,1)$. It should be understood that we can apply many functions to transform a random variable. In this case the function we are applying is the cdf itself. This special type of transformation is known as the Probability Integral Transform (PIT). We will discuss the PIT later in this course as a way to simulate values from candidate distributions to fit to a data set. The proof for this statement is as follows

Define the random variable $U \sim U(0,1)$ having cdf $F_U(u) = P(U \leq u) = u$. If we apply the transformation $U = F_T(T)$ where T is any continuous random variable then

$$
\begin{aligned}
P(U \leq u) &= P(F_T(t) \leq u) \\
&= P(T \leq F_T^{-1}(u)) \\
&= F_T(F_T^{-1}(u)) \\
&= u \qquad\qquad q.e.d.
\end{aligned}
$$

Now, we have established that $F_T(T) \sim U(0,1)$, therefore we define $U = F_T(T)$. We want to find the distribution of $V = H_T(T)$. We can accomplish this using either the cdf or the pdf of V, let's start with the cdf:

Because of the relationship between the cdf and the cumulative hazard we know $U = 1 - \exp\{-V\}$. Therefore $V = -\log(1-U), 0 < V < \infty$. The cdf of V may then be expressed as

$$
\begin{aligned}
P(V \leq v) &= P(-\log(1-U) \leq v) \\
&= P(1-U > \exp(-v)) \\
&= P(U \leq 1 - \exp(-v)) \\
&= F_U(1 - \exp(-v) \\
&= 1 - \exp(-V)
\end{aligned}
$$

Which is the cdf of the Exp(1) distribution

For the pdf case, we note that the Jacobian J of the transformation is

$$
J = \left| \frac{d}{dV} U \right| = \exp\{-V\}
$$

and $f_V(v) = f_U(1 - \exp(-v))J = \exp\{-V\}$ for $0 < V < \infty$ which is the pdf of the Exp(1) distribution.