HW 3-1 (Posted on March 18, 2017) **Name:**

HW 3.1. (Continuous Distribution) Suppose that X has the pdf

$$f(x) = \begin{cases} ax, & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of a (use the requirement $\int_{-\infty}^{\infty} f(x)dx = 1$).

(b) Calculate $P(X < 0.3)$.

(c) Calculate $P(0.3 \leq X < 0.8)$.

(d) Find the mean and variance of X.

(e) If we define $Y = 3X$, calculate the mean and variance of Y.

Sol.

(a) Since $1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^1 axdx = a/2$, then we have $a = 2$ and

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(b) $P(X < 0.3) = \int_{-\infty}^{0.3} f(x)dx = \int_0^{0.3} 2xdx = 0.09$.

(c) $P(0.3 < X < 0.8) = \int_{0.3}^{0.8} f(x)dx = \int_{0.3}^{0.8} 2xdx = 0.55$.

(d) $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x \cdot 2xdx = 2/3$.
$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 x^2 \cdot 2xdx = 2/4 = 1/2$.
Then we have $V(X) = E(X^2) - [E(X)]^2 = 1/2 - (2/3)^2 = 1/18$.

(e) $E(Y) = E(3X) = 3E(X) = 3(2/3) = 2$.
$V(Y) = V(3X) = 3^2 V(X) = 9(1/18) = 1/2$.

HW 3.2. (Exponential Distribution) Web crawlers need to estimate the frequency of changes to Web sites to maintain a current index for Web searches. Assume that the changes to a Web site follow a Poisson process with a mean of 3.5 days. Then the time to the next change occurs has exponential distribution with waiting parameter $\lambda = 1/3.5$.

(a) What is the probability that the next change occurs in less than 2.0 days?

(b) What is the probability that the time until the next change is greater 7.0 days?

(c) What is the time of the next change that is exceeded with probability 90%?

(d) What is the probability that the next change occurs in less than 10.0 days, given that it has not yet occurred after 3.0 days?

Sol. Define T be the time to the next change occurs, then

$$X \sim \text{Exp}(\lambda = 1/3.5)$$

(a) $P(T < 2.0) = 1 - e^{-(1/3.5)2} = 1 - 0.5647 = 0.4353$.

(b) $P(T > 7.0) = e^{-(1/3.5)7} = 0.1353$.

(c) Find t such that $0.9 = P(T > t) = e^{-3.5t}$. Then we have $t = -\log(0.9) \times 3.5 = 0.3688$.

(d) Note that $P(T < 10|T > 3) = 1 - P(T > 10|T > 3)$. According to memoryless properties, $P(T > 10|T > 3) = P(T > 7) = 0.1353$, then we have $P(T < 10|T > 3) = 1 - P(T > 7) = 0.8647$. Or, we can calculate the probability directly.

$$\begin{aligned} P(T < 10|T > 3) &= \frac{P(T < 10 \text{ and } T > 3)}{P(T > 3)} = \frac{P(3 < T < 10)}{P(T > 3)} = \frac{P(T < 10) - P(T \leq 3)}{e^{-(1/3.5)3}} \\ &= \frac{(1 - e^{-(1/3.5)10}) - (1 - e^{-(1/3.5)3})}{e^{-(1/3.5)3}} = . \end{aligned}$$

HW 3.3 (Exponential Distribution) The length of stay at a specific emergency department in a hospital in Phoenix, Arizona had a mean of 4.6 hours. Assume that the length of stay is exponentially distribution.

(a) What is the standard deviation of the length of stay?

(b) What is the probability of a length of stay of more than 10 hours.

(c) What length of stay is exceeded by 25% of the visits?

Sol. Let T be the length of stay at a specific emergency department in a hospital in Phoenix. Then we have

$$X \sim \text{Exp}(\lambda = 1/4.6).$$

(a) $Sd(T) = \sqrt{Var(T)} = \sqrt{1/\lambda^2} = 1/\lambda = 4.6.$

(b) $P(T > 10) = e^{(-1/4.6)(10)} = 0.1137.$

(c) Find t such that $0.25 = P(T > t) = e^{(-1/4.6)t}$. Then we have $t = [\log 0.25] \times (-4.6) = 6.3770.$

HW 3.4. (Gamma distribution) Errors caused by contamination on optical disks occur at the rate of one error every 10^5 bits. Assume that the errors follow a Poisson distribution.

(a) What is the mean number of bits until five errors occur?

(b) What is the standard deviation of the number of bits from the fifth error to ninth error?

(c) The error-correcting code might be ineffective if there are three or more errors within 10^5 bits. What is the probability of this event?

Sol. Let X be the number of errors in every 10^5 bits. Here we treat 10^5 bits as a unit of time. Then we have

$$X \sim \text{ Poisson } (\lambda = 1),$$

since the rate of error is one for every 10^5 bits.
(a) Let T be the unit of time until five errors occur. According to relation between Poisson and Gamma distribution, we have $T \sim$Gamma$(\alpha = 5, \lambda = 1)$. Then $E(T) = \alpha/\lambda = 5/1 = 5$. In other words, the mean is 5 units, i.e., 5×10^5 bits.

(b) Let W be the units of time from fifth to ninth error occurs. According to relation between Poisson and Gamma distribution, we have $W \sim$ Gamma$(\alpha = 9 - 5, \lambda = 1)$, i.e., $W \sim$ Gamma$(\alpha = 4, \lambda = 1)$. Then $Sd(W) = \sqrt{Var(W)} = \sqrt{4/1^2} = 2$. In other words, the standard deviation is 2 units, i.e., 2×10^5 bits.

(c) $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2)$. By using R, we have $1 - P(X \leq 2) = 0.0803$.

HW 3.5 (Normal distribution) If X is normally distributed with a mean of 6 and a standard deviation of 2.

(a) Find the followings probabilities.

- $P(2 < X < 10) = \texttt{normcdf}(2, 10, 6, 2) = 0.9545$.
- $P(X > 8) = \texttt{normcdf}(8, 10^{99}, 6, 2) = 0.1587$.
- $P(X < 0) = \texttt{normcdf}(-10^{99}, 0, 6, 2) = 0.0013$.
- $P(3 \leq X < 9) = \texttt{normcdf}(3, 9, 6, 2) = 0.8664$.
- $P(X \geq 7) = \texttt{normcdf}(7, 10^{99}, 6, 2) = 0.3085$.
- $P(X < 1) = \texttt{normcdf}(-10^{99}, 1, 6, 2) = 0.0062$.

(b) Determine the value for x that solves each of the following:

- $P(X > x) = 0.5$. $x = \texttt{invNorm}(1 - 0.5, 6, 2) = 6$.
- $P(X > x) = 0.95$. $x = \texttt{invNorm}(1 - 0.95, 6, 2) = 2.7103$.
- $P(x < X < 9) = 0.2$
- $P(2 < X < x) = 0.95$
- $P(-x < X - 6 < x) = 0.99$

Sol.

(b)3. Since $0.2 = P(x < X < 9) = P(X < 9) - P(X \leq x)$, then we have $P(X \leq x) = P(X < 9) - 0.2 = \texttt{normalcdf}(-10^{99}, 9, 6, 2) - 0.2 = 0.9332 - 0.2 = 0.7332$. Then $x = \texttt{invNorm}(0.7332, 6, 2) = 7.2445$.

(b)4. Since $0.95 = P(2 < X < x) = P(X < x) - P(X \leq 2)$, then we have $P(X < x) = 0.95 + P(X \leq 2) = 0.95 + \texttt{normalcdf}(-10^{99}, 2, 6, 2) = 0.95 + 0.0227 = 0.9727$. Then $x = \texttt{invNorm}(0.9727, 6, 2) = 9.8441$.

(b)5. Since $0.99 = P(-x < X - 6 < x) = 1 - 2P(X - 6 > x)$(symmetric about mean 6), then $P(X - 6 > x) = (1 - 0.99)/2 = 0.005$. Let $C = x + 6$, we solve C such that $P(X > c) = P(X > x + 6) = P(X - 6 > x) = 0.005$ and obtain $C = \texttt{invNorm}(1 - 0.005, 6, 2) = 11.1516$ and $x = C - 6 = 11.1516 - 6 = 5.1516$.

HW 3.6 (Normal distribution) The fill volume of an automated filling machine used for filling cans of carbonated beverage is normally distributed with a mean of 12.4 fluid ounces and a standard deviation of 0.1 fluid ounce.

(a) What is the probability that a fill volume is less than 12 fluid ounces?

(b) If all cans less than 12.1 or more than 12.6 ounces are scrapped, what proportion of cans is scrapped.

(c) Suppose the variance is unknown, and we know the probability that a fill volume exceeds 13 is 0.05, find the variance.

Sol. Let X is the volume of an automated filling machine used for filling cans of carbonated beverage, then we have $X \sim N(\mu = 12.4, \sigma^2 = (0.1)^2)$.
(a) $P(X < 12) = \texttt{normalcdf}(-10^{99}, 12, 12.4, 0.1) = 3.1671 \times 10^{-5}$.

(b) $P(X < 12.1 \text{ or } X > 12.6) = 1 - P(12.1 \leq X \leq 12.6) = \texttt{normalcdf}(12.1, 12.6, 12.4, 0.1) = 0.9759$.

(c) Here we solve

$$0.05 = P(X > 13) = P\left(\frac{X - 12.4}{\sigma} > \frac{13 - 12.4}{\sigma}\right) = P\left(Z > \frac{13 - 12.4}{\sigma}\right) = P\left(Z > \frac{0.6}{\sigma}\right).$$

where Z has standard normal distribution, i.e., $Z \sim N(0, 1)$. Define $c = \frac{0.6}{\sigma}$, then we fine c such that $0.05 = P(Z > c)$. By using `invNorm` function, we can find $c = \texttt{invNorm}(1 - 0.05, 0, 1) = 1.645$. Then we have $\sigma = \frac{0.6}{c} = \frac{0.6}{1.645} = 0.3647$ and $\sigma^2 = (0.3647)^2 = 0.1330$.

HW 3.7 (Weibull distribution)

An article in *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* ["Understanding Web Browsing Behaviors Through Weibull Analysis of Dwell Time" (2010, p.379-386)] proposed that a Weibull distribution can be used to model Web page dwell time (the length of time a Web visitor spends on a Web page). For a specific Web page, the shape and scale parameters are 1 and 300 seconds, respectively. Determine the following:

(a) Mean and variance of dwell time.

(b) Probability that a Web user spends more than four minutes on this Web page.

(c) Dwell time exceeded with probability 0.25.

Sol. Let T be the Web page dwell time. Then $T \sim$Weibull$(\beta = 1, \eta = 300)$.

(a) $E(T) = \eta\Gamma\left(1 + \frac{1}{\beta}\right) = 300\,\Gamma(1+1) = 300.$

$$Var(T) = \eta^2 \left\{\Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2\right\} = 300^2 \left\{\Gamma\left(1+2\right) - [1]^2\right\} = 90000(2-1) = 90000.$$

(b) $P(T > 4 \times 60) = e^{-\left(\frac{4\times 60}{\eta}\right)^{\beta}} = e^{-\left(\frac{240}{300}\right)^{1}} = 0.4493.$

(c) $0.25 = P(T > t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}} = e^{-\left(\frac{t}{300}\right)^{1}}$. Then we solve $t = -300\,\log(0.25) = 415.8883$.

HW 3.8 (Weibull distribution) A shock absorber is a suspension component that controls the up-and-down motion of a vehicle's wheels. The following data are $n = 30$ distance (in km) to failure for a certain brand of chock absorber under "extreme" driving conditions.

6700	6950	9120	9660	11310	11850	11880	12140	12200	13330
13470	14040	14300	17520	17890	18450	18960	18980	19410	20100
20100	20320	20900	22700	26510	27410	27490	27890	28100	30050

Under a Weibull assumption for

$$T = \text{distance (in km) to failure},$$

we calculate the maximum likelihood estimate $\hat{\beta} = 2.96$ and $\hat{\eta} = 19899.53$ for the data above by `R`. Use these values (along with the Weibull assumption) to answer the following questions.)

(a) Calculate $P(T > 15000)$. Interpret what is this probability means in words.

(b) Find the 25th percentile of the distance to failure distribution. Interpret in words.

(c) Find the hazard function $h_T(t)$ and explain, in plain English, what information of this function reveals in this example?

(d) Plot the estimated hazard function for T. Explain, in plain English, what information this graph reveals.

Sol. (a) Using the estimated Weibull($\hat{\beta} = 2.9584, \hat{\eta} = 19899.53$) distribution as a model, we have $P(T > 15000) = 0.6483$ which is the probability a shock absorber will "survive" past 15000 km.

(b) Using `R`, the 25th percentile is 13060.03. 25% of the shock absorbers will fail before 13060.03 km.

(c) Since $\hat{\beta} = 2.9584 > 1$, the hazard function is increasing and the population of shock absorbers should be getting weaker overtime. Furthermore, the rate of failure increases at an increasing rate.

(d)

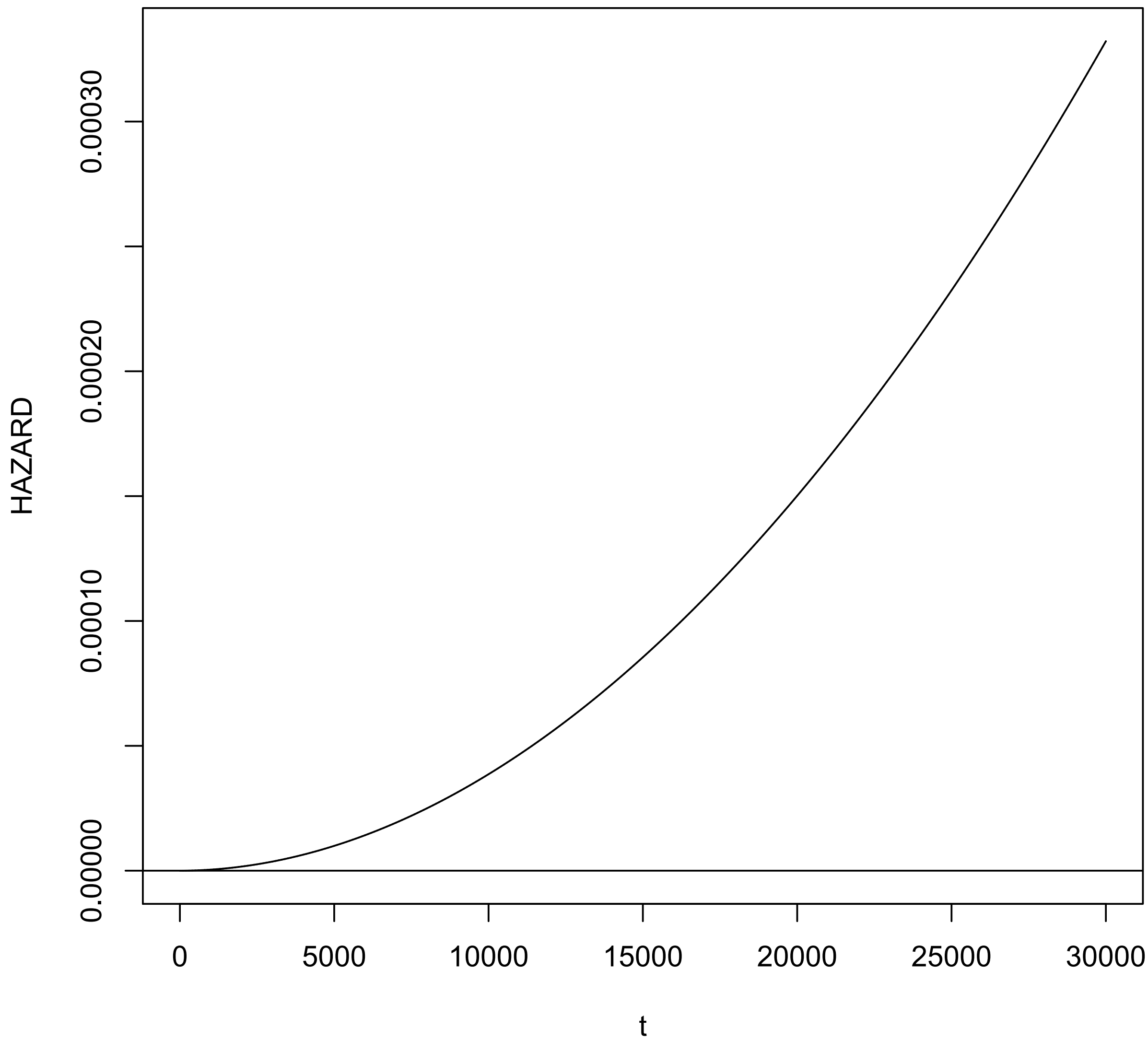

Figure 0.0.1: Hazard function of Shock data