Experimental Statistics for Engineers I

Kamala Dadashova

Problem 1: The CDF for a RV is

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{2} + \frac{3}{32}(4x - \frac{x^3}{3}) & -2 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

(a) Find $P(X < 0)$.

$P(X < 0) = F(0) = 0.5$

(b) Find $P(-1 < X < 1)$

$P(-1 < X < 1) = F(1) - F(-1) = \frac{1}{2} + \frac{3}{32}(4 - \frac{1^3}{3}) - (\frac{1}{2} + \frac{3}{32}(-4 - \frac{-1^3}{3})) = \frac{11}{16}$

(c) Find $P(0.5 < X)$.

$P(0.5 < X) = 1 - P(X \leq 0.5) = 1 - F(0.5) = 1 - 0.6836 = 0.3164$

(d) Find $f(x)$ the pdf of X.

Note that $f(x) = F'(x) => (\frac{1}{2} + \frac{3}{32}(4x - \frac{x^3}{3}))' = \frac{3}{32}(4 - x^2)$

$$f(x) = \begin{cases} 0 & otherwise \\ \frac{3}{32}(4 - x^2) & -2 \leq x < 2 \end{cases}$$

(e) Find $E(X)$

$$\mu_x = E(X) = \int_{-\infty}^{\infty} xf(x)\ dx = \int_{-2}^{2} \frac{3}{32}(4 - x^2)x\ dx = 0$$

Problem 2: A college professor never finishes his lecture before the end of the hour and always finishes his lectures within 2 min after the hour. Let X = the time that elapses between the end of the hour and the end of the lecture and suppose the pdf of X is

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq 2 \\ 0 & otherwise \end{cases}$$

(a) Find the value of k and draw the corresponding density curve. [Hint: Total area under the graph of f (x) is 1.]

$$1 = \int_{-\infty}^{\infty} f(x)\ dx$$
$$1 = \int_{0}^{2} kx^2\ dx$$
$$k = \frac{3}{8}$$

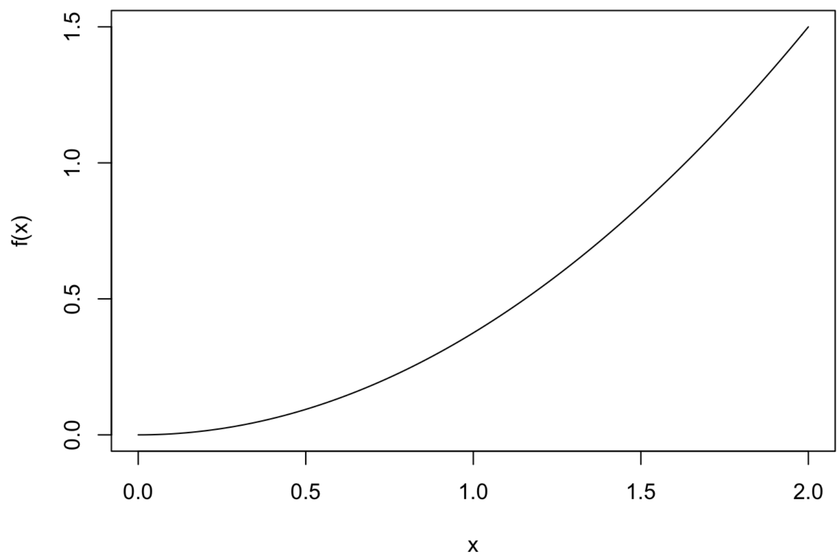

Figure 1: Density curve

(b) What is the probability that the lecture ends within 1 min of the end of the hour?

$$P(0 \leq X \leq 1) = \int_{0}^{1} \frac{3}{8}x^2\ dx = \frac{1}{8}$$

(c) What is the probability that the lecture continues beyond the hour for between 60 and 90 sec?

$$P(1 \leq X \leq 1.5) = \int_{1}^{1.5} \frac{3}{8}x^2\ dx = \frac{19}{64}$$

(d) What is the probability that the lecture continues for at least 90 sec beyond the end of the hour?

$$P(X \geq 1.5) = 1 - \int_{0}^{1.5} \frac{3}{8}x^2\ dx = 0.578125$$

Problem 3: Consider the pdf for total waiting time Y for two buses

$$f(y) = \begin{cases} \frac{1}{25}y & 0 \le y < 5 \\ \frac{2}{5} - \frac{1}{25}y & 5 \le y \le 10 \\ 0 & otherwise \end{cases}$$

(a) Compute and sketch the cdf of Y.

For interval $0 \le y < 5$,

$$F(y) = \int_0^y f(t)\ dt = \int_0^y \frac{1}{25}t\ dt = \frac{y^2}{50}$$

For interval $5 \le y \le 10$

$$F(y) = \int_0^y f(t)\ dt = \int_0^5 \frac{1}{25}t\ dt + \int_5^y (\frac{2}{5} - \frac{1}{25}t)\ dt = \frac{2}{5}y - \frac{y^2}{50} - 1$$

Thus,

$$F(y) = \begin{cases} 0 & y < 0 \\ \frac{y^2}{50} & 0 \le y < 5 \\ \frac{2}{5}y - \frac{y^2}{50} - 1 & 5 \le y \le 10 \\ 1 & y > 10 \end{cases}$$

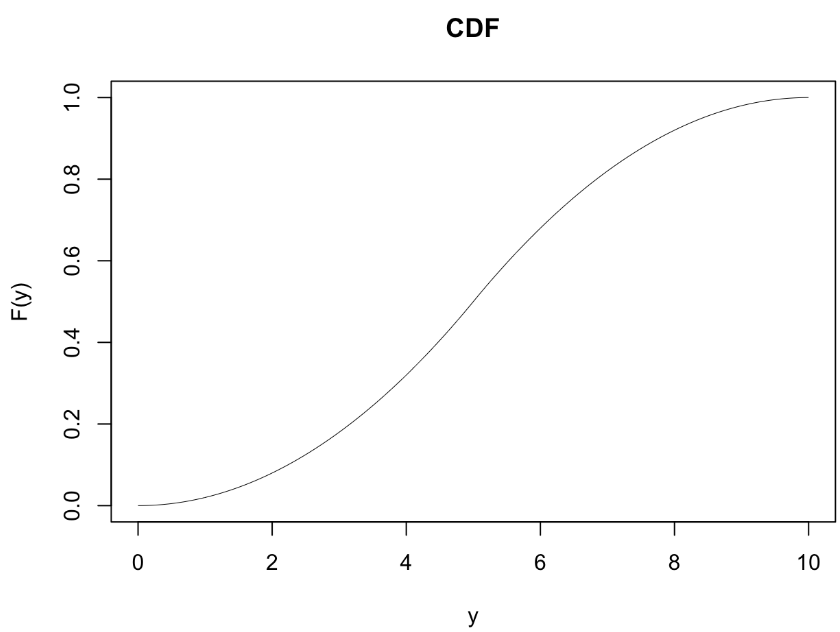

Figure 2

(b) Obtain an expression for the (100p)th percentile.

For $0 \le p < 5$, $p = F(y_p)$, $y_p = \sqrt{(50p)}$

For $5 \le p \le 10$, $p = F(y_p)$, $p = \frac{2}{5}y_p - \frac{y_p^2}{50} - 1$,$y_p = 10 - 5\sqrt{2(1-p)}$ note that we take this root because another root is bigger than 10.

(c) Compute E(Y) and V(Y). How do these compare with the expected waiting time and variance for a single bus when the time is uniformly distributed on [0, 5]?

$$\mu = E(Y) = \int_{-\infty}^{\infty} y f(y)\, dy = \int_0^5 \frac{1}{25} y^2\, dy + \int_5^{10} y(\frac{2}{5}y - \frac{y^2}{50} - 1)\, dy = 5$$

$$V(Y) = \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\, dy =_0^5 \frac{1}{25} y\,(y-5)^2\, dy + \int_5^{10} \left(\frac{2}{5} - \frac{1}{25}y\right)(y-5)^2\, dy \ = \frac{25}{12} \approx 4.1667$$

We know that if $Y \sim U(a,b)$, then

$$E(Y) = \frac{a+b}{2}$$

$$V(Y) = \frac{(b-a)^2}{12}$$

Then, for $Y \sim U(0,5)$, we have

$$E(Y) = \frac{0+5}{2} = 2.5$$

$$V(Y) = \frac{(5-0)^2}{12} = \frac{25}{12}$$

As we observe although expectation values for both cases are different, their variances are same.

Problem 4: In a road-paving process, asphalt mix is delivered to the hopper of the paver by trucks that haul the material from the batching plant. The article “Modeling of Simultaneously Continuous and Stochastic Construction Activities for Simulation” (J. of Construction Engr. and Mgmnt., 2013: 1037–1045) proposed a normal distribution with mean value 8.46 min and standard deviation .913 min for the rv X = truck haul time.

(a) What is the probability that haul time will be at least 10 min? Will exceed 10 min?

$$1 - pnorm(q = 10, mean = 8.46, sd = 0.913) = 0.0458$$

We know that since X is continuous, $P(X > 10) = P(X \ge 10) = 0.0458$

(b) What is the probability that haul time will exceed 15 min?

$$P(X > 15) = P(Z > \frac{15 - 8.46}{0.913}) = 0$$

$$pnorm(q = 15, mean = 8.46, sd = 0.913, lower.tail = FALSE) \approx 0$$

(c) What is the probability that haul time will be between 8 and 10 min?

$$P(8 \leq X \leq 10) = P(\frac{8-8.46}{0.913} \leq Z \leq \frac{10-8.46}{0.913}) = 0.6469848$$

$$pnorm(10, 8.46, 0.913) - pnorm(8, 8.46, 0.913)$$

(d) What value c is such that 98% of all haul times are in the interval from 8.46-c to 8.46 +c?

98% will be in the middle of the distribution. Remaining 2% will be equally distributed to left and right side of distribution.

On the left,

$$\mu - c = 8.46 - c = qnorm(p = .01, mean = 8.46, sd = 0.913) = 6.34$$
$$c = 2.12$$

On the right,

$$\mu + c = 8.46 + c = qnorm(p = .99, mean = 8.46, sd = 0.913) = 10.58$$
$$c = 2.12$$

(e) If four haul times are independently selected, what is the probability that at least one of them exceeds 10 min?

$$P(X > 10) = 1 - P(X \leq 10)$$

With R,

$$P(X \leq 10) = pnorm(q = 10, mean = 8.46, sd = 0.913) = 0.9542.$$

Then, we have

$$1 - (0.9542)^4 = 0.171$$

Problem 5:

(a) Show that if X has a normal distribution with parameters μ and σ, then $Y = aX + b$ (a linear function of X) also has a normal distribution. What are the parameters of the distribution of Y [i.e., E(Y) and V(Y)]?

Consider $a > 0$. The cumulative function with respect to Y,

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a})$$

We know that

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(\frac{y-b}{a}) = \frac{1}{a}f_X(\frac{y-b}{a})$$

It is given that X has a normal distribution,

$$f_Y(y) = \frac{1}{a}\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{(-\frac{y-b}{a}-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}a\sigma}e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}}$$

It reminds us normal distribution with $E(Y) = a\mu + b$ and $V(Y) = a^2\sigma^2$. Without losing generality, we can do the same thing when $a < 0$.

(b) If, when measured in $8C$, temperature is normally distributed with mean 115 and standard deviation 2, what can be said about the distribution of temperature measured in $8F$?

We know that the conversion formula is $F = 1.8C + 32$. Given that C is normally distributed. Using part (a), we can say that F is also normal distribution with $mu = 1.8 * 115 + 32 = 239F$ and $variance = 1.8^2 2^2 = 12.96F$ i.e. $standarddeviation = 3.6F$

Problem 6: A consumer is trying to decide between two long-distance calling plans. The first one charges a flat rate of 10¢ per minute, whereas the second charges a flat rate of 99¢ for calls up to 20 minutes in duration and then 10¢ for each additional minute exceeding 20 (assume that calls lasting a noninteger number of minutes are charged proportionately to a whole-minute's charge). Suppose the consumer's distribution of call duration is exponential with parameter l.

(a) Explain intuitively how the choice of calling plan should depend on what the expected call duration is.

Two cases might happen. First, if we call for short time then the first plan makes more sense to use it. The short time is any minutes between 0 and 10, exclusive). But if talk minutes are greater than and equal to 10 then the second one is good to use.

(b) Which plan is better if expected call duration is 10 minutes? 15 minutes?

Let us write the equation for the given information.

$$h_1(X) = 10X$$

$$h_2(X) = \begin{cases} 99 & 0 \le X \le 20 \\ 99 + 10(X - 20) & X > 20 \end{cases}$$

We know that $E(X) = \mu = \frac{1}{\lambda}$ for exponential distribution. Then,

$$E[h_1(X)] = 10E[X] = 10\mu$$

$$E[h_2(X)] = 99 + 10\int_{20}^{\infty}(x-2)\lambda e^{-\lambda x}\,dx = 99 + \frac{10}{\lambda}e^{-20\lambda}$$

At $\mu = 10$, $E[h_1(X)] = \$1$, $E[h_2(X)] = \$99 + 100e = \$1.13$ At $\mu = 15$, $E[h_1(X)] = \$1.50$, $E[h_2(X)] = \$99 + 150e^{\frac{-4}{3}} = \$1.39$

Problem 7: Suppose that when a transistor of a certain type is subjected to an accelerated life test, the lifetime X (in weeks) has a gamma distribution with mean 24 weeks and standard deviation 12 weeks.

(a) What is the probability that a transistor will last between 12 and 24 weeks?

For gamma distribution, we know that

$$E(X) = \alpha\beta$$

$$V(X) = \alpha\beta^2$$

Thus, we have

$$\alpha\beta = 24$$

$$\alpha\beta^2 = 12^2$$

By solving this system, we get $\alpha = 4, \beta = 6$. Then we use R to find probability,

$$pgamma(q = 24, shape = 4, scale = 6) - pgamma(q = 12, shape = 4, scale = 6) = 0.424$$

(b) What is the probability that a transistor will last at most 24 weeks? Is the median of the lifetime distribution less than 24? Why or why not?

$$P(X \leq 24) = pgamma(q = 24, shape = 4, scale = 6) = 0.567$$

Although mean is 24, the median is less than 24 since $P(x \leq \eta) = 0.5$ (so it is less than 0.567). Note that it is positive skew gamma distribution.

(c) What is the 99th percentile of the lifetime distribution?

$$qgamma(p = 0.99, shape = 4, scale = 6) = 60.27071 \approx 60\ weeks$$

(d) Suppose the test will actually be terminated after t weeks. What value of t is such that only .5% of all transistors would still be operating at termination?

We want to find t such that $P(X > t) = .005$ or equivalently $P(X \leq t) = .995$. Using R,

$$qgamma(p = 0.995, shape = 4, scale = 6) = 65.86486 \approx 66\ weeks$$

Problem 8: Stress is applied to a 20-in. steel bar that is clamped in a fixed position at each end. Let Y=the distance from the left end at which the bar snaps. Suppose Y/20 has a standard beta distribution with $E(Y) = 10$ and $V(Y) = \frac{100}{7}$.

(a) What are the parameters of the relevant standard beta distribution?

The expectation of $\frac{Y}{20}$ is

$$E(\frac{Y}{20}) = \frac{1}{20}E(Y) = \frac{1}{20}10 = \frac{1}{2}$$

The variance of $\frac{Y}{20}$ is

$$V(\frac{Y}{20}) = \frac{1}{400}V(Y) = \frac{1}{400}\frac{100}{7} = \frac{1}{28}$$

$$E(\frac{Y}{20}) = \frac{\alpha}{\alpha + \beta}$$

$$V(\frac{Y}{20}) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
$$\frac{1}{2} = \frac{\alpha}{\alpha+\beta}$$
$$\alpha = \beta$$
$$\frac{1}{28} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

By solving this equation, we get $\alpha = 3$ and $\beta = 3$.

(b) Compute $P(8 \leq Y \leq 12)$.

We want to find
$$P(8 \leq Y \leq 12) = P(\frac{8}{20} \leq \frac{Y}{20} \leq \frac{12}{20})$$

We used R to compute this with the following code,

$pbeta(q = 12/20, shape1 = 3, shape2 = 3) - pbeta(q = 8/20, shape1 = 3, shape2 = 3) = 0.36512$

(c) Compute the probability that the bar snaps more than 2 in. from where you expect it to.

The probability that bar snaps more than 2 inches from the expected value(10) is 12,
$$P(Y > 12) = P(\frac{Y}{20} > \frac{12}{20})$$

With R,
$$pbeta(12/20, shape1 = 3, shape2 = 3, lower.tail = F) = 0.31744$$

or equivalently,
$$1 - pbeta(12/20, shape1 = 3, shape2 = 3) = 0.31744$$