

Homework 4

Problem 1. The Weibull cumulative distribution function is

$$F(x) = 1 - e^{-(x/\alpha)^\beta}, \quad x \geq 0, \quad \alpha > 0, \quad \beta > 0$$

(a) Find the density function.

(b) Show that if W follows a Weibull distribution, then $X = (W/\alpha)^\beta$ follows an exponential distribution.

(c) How could Weibull random variables be generated from a uniform random number generator?

Proof. (a) The density function is the derivative of $F(x)$. That is

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx} \left(1 - e^{-(x/\alpha)^\beta} \right) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha} \right)^{\beta-1} e^{-(x/\alpha)^\beta}.$$

(b) Let $X = g(W) = (W/\alpha)^\beta$. Then $g^{-1}(x) = (\alpha/\beta) \log x$. We know

$$\begin{aligned} f_X(x) &= f_W \left(\left(\frac{\alpha}{\beta} \right) \log x \right) \left| \frac{1}{\frac{\beta}{\alpha} \left(\frac{g^{-1}(x)}{\alpha} \right)^{\beta-1}} \right| \\ &= \frac{\beta}{\alpha} \left(\frac{1}{\alpha} \left(\frac{\alpha}{\beta} \right) \log x \right)^{\beta-1} \exp \left(- \left(\frac{1}{\alpha} \left(\frac{\alpha}{\beta} \right) \log x \right)^\beta \right) \left| \frac{\alpha}{\beta} \left(\frac{\log x}{\beta} \right)^{1-\beta} \right| \\ &= \exp \left(- \left(\frac{\log x}{\beta} \right)^\beta \right) \\ &= \exp \left(- \exp \left(\log \left(\frac{\log x}{\beta} \right)^\beta \right) \right) \\ &= \exp \left(- \exp \left(\beta \log \left(\frac{\log x}{\beta} \right) \right) \right) \\ &= \exp \left(- e^\beta \exp \left(\log \left(\frac{\log x}{\beta} \right) \right) \right) \\ &= \exp(-\exp(\log x)) \\ &= e^{-x}. \end{aligned}$$

(c) We know that if we apply F^{-1} to a uniform distribution, the resulting cdf will be F . But note that

$$F^{-1}(y) = a \left(\log \left(\frac{1}{1-y} \right) \right)^{\frac{1}{\beta}},$$

so we can use this applied to a uniform random number generator to get a Weibull distribution. \square

Problem 2. An urn contains p black balls, q white balls and r red balls; and n balls chosen without replacement.

(a) Find the joint distribution of the numbers of black, white and red balls in the sample.

(b) Find the joint distribution of the numbers of black and white balls in the sample.

(c) Find the marginal distribution of the number of white balls in the sample.

(a) Let B , W and R be the distributions of black, white and red balls chosen, respectively. Let $m = p + q + r$. Then

$$p(n_1, n_2, n_3) = P(B = n_1, W = n_2, R = n_3) = \frac{\binom{p}{n_1} \binom{m-p}{n-n_1} \binom{q}{n_2} \binom{m-p-q}{n-n_1-n_2}}{\binom{m}{n}} = \frac{\binom{p}{n_1} \binom{q}{n_2} \binom{r}{n_3}}{\binom{m}{n}}$$

where we've used that $n = n_1 + n_2 + n_3$.

(b) Note that in part (a), the probability of choosing n_3 red balls was already determined by the number of black and white balls chosen since $n_3 = n - n_1 - n_2$. Thus, we must have the same probability distribution for black and white as black and white and red. Therefore

$$p(n_1, n_2) = P(B = n_1, W = n_2) = \frac{\binom{p}{n_1} \binom{q}{n_2} \binom{r}{n_3}}{\binom{m}{n}}.$$

(c) To find the marginal probability we can take the joint distribution from part (b) and sum over all possible values for B . This is

$$p(k) = \sum_{j=0}^{n-k} \frac{\binom{q}{k} \binom{p}{j} \binom{r}{n-k-j}}{\binom{m}{n}}.$$

Problem 3. A point is chosen randomly in the interior of an ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Find the marginal densities of the x and y coordinates of the point.

The area of the ellipse is πab , so

$$f(x, y) = \begin{cases} \frac{1}{\pi ab} & \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

We can now find the marginal density of the x coordinate as

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{\pi ab} \int_{-b\sqrt{1-x^2/a^2}}^{b\sqrt{1-x^2/a^2}} dy = \frac{2}{a\pi} \sqrt{1 - \frac{x^2}{a^2}}$$

where $|x| \leq a$. By symmetry,

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

where $|y| \leq b$.

Problem 4. Let X and Y and the joint density

$$f(x, y) = \frac{6}{7}(x+y)^2, \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1$$

(a) By integrating over the appropriate regions, find (i) $P(X > Y)$, (ii) $P(X + Y \leq 1)$, (iii) $P(X \leq \frac{1}{2})$.

(b) Find the marginal densities of X and Y .

(c) Find the two conditional densities.

(a) (i) We integrate over the set $\{(x, y) \mid 0 \leq y \leq x \leq 1\}$. This is

$$P(X > Y) = \frac{6}{7} \int_0^1 \int_0^x (x^2 + 2xy + y^2) dy dx = \frac{6}{7} \int_0^1 \frac{7}{3} x^3 dx = \frac{1}{4}.$$

(ii) We integrate over the set $\{(x, y) \mid 0 \leq y \leq 1 - x \leq 1\}$. This is

$$P(X + Y \leq 1) = \frac{6}{7} \int_0^1 \int_0^{1-x} (x^2 + 2xy + y^2) dy dx = \frac{6}{7} \int_0^1 \left(\frac{1}{3} - \frac{x^3}{3} \right) dx = \frac{3}{14}.$$

(iii) This is the marginal cdf of X evaluated at $\frac{1}{2}$. This is

$$P\left(X \leq \frac{1}{2}\right) = F_X\left(\frac{1}{2}\right) = \lim_{y \rightarrow \infty} F(x, y) = \frac{6}{7} \int_0^{\frac{1}{2}} \int_0^1 (x^2 + 2xy + y^2) dy dx = \frac{6}{7} \int_0^{\frac{1}{2}} x^2 + x + \frac{1}{3} dx = \frac{2}{7}.$$

(b) The marginal density of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{6}{7} \int_0^1 (x^2 + 2xy + y^2) dy = \frac{6}{7} \left(x^2 + x + \frac{1}{3} \right).$$

The marginal density of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \frac{6}{7} \int_0^1 (x^2 + 2xy + y^2) dx = \frac{6}{7} \left(y^2 + y + \frac{1}{3} \right).$$

(c) We know $f_{XY}(x, y) = \frac{6}{7}(x + y)^2$, $f_X(x) = \frac{6}{7}(x^2 + x + \frac{1}{3})$ and $f_Y(y) = \frac{6}{7}(y^2 + y + \frac{1}{3})$. Then

$$f_{X|Y}(x | y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{(x + y)^2}{y^2 + y + \frac{1}{3}}$$

and

$$f_{Y|X}(y | x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{(x + y)^2}{x^2 + x + \frac{1}{3}}.$$

Problem 5. Let

$$f(x, y) = c(x^2 - y^2)e^{-x}, \quad 0 \leq x \leq \infty, \quad -x \leq y \leq x$$

(a) Find c .

(b) Find the marginal densities.

(c) Find the conditional densities.

(a) We can find c by integrating $f(x, y)$ over all space as

$$1 = c \int_0^{\infty} \int_{-x}^x (x^2 - y^2) e^{-x} dy dx = \frac{4c}{3} \int_0^{\infty} x^3 e^{-x} dx = 8c$$

so $c = 1/8$.

(b) We have

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{8} \int_{-x}^x (x^2 - y^2) e^{-x} dy = \frac{x^3 e^{-x}}{6}.$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \frac{1}{8} \int_0^{\infty} (x^2 - y^2) e^{-x} dx = \frac{2 - y^2}{8}.$$

(c) Using part (b) we have

$$f_{X|Y}(x | y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{(x^2 - y^2)e^{-x}}{2 - y^2}$$

and

$$f_{Y|X}(y | x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{3}{4} \frac{x^2 - y^2}{x^3}.$$

Problem 6. An instrument is used to measure very small concentrations, X , of a certain chemical in soil samples. Suppose that the values of X in those soils in which the chemical is present is modeled as a random variable with density function $f(x)$. The assay of a soil reports a concentration only if the chemical is first determined to be present. At very low concentrations, however, the chemical may fail to be detected, even if it is present. This phenomenon is modeled by assuming that if the concentration is x , the chemical is detected with probability $R(x)$. Let Y denote the concentration of a chemical in a soil in which it has been determined to be present. Show that the density function of Y is

$$g(y) = \frac{R(y)f(y)}{\int_0^\infty R(x)f(x)dx}.$$

Proof. The two densities $f(x)$ and $R(x)$ make a joint distribution with density $R(x)f(x)$ which represents the probability of the instrument measuring the chemical. Then the marginal distribution density for R is $\int_0^\infty R(x)f(x)dx$. Note that Y is defined to be the conditional distribution of X given R , so by definition, the density function for Y is the joint density function $R(y)f(y)$ divided by the marginal distribution for R . Namely,

$$g(y) = \frac{R(y)f(y)}{\int_0^\infty R(x)f(x)dx}.$$

□

Problem 7. Show that the number of iterations necessary to generate a random variable using the rejection method is a geometric random variable, and evaluate the parameter of the geometric frequency function. Show that in order to keep the number of iterations small, $M(x)$ should be chosen to be close to $f(x)$.

Proof. The method can be described as picking a point in the region defined by $M(x)$, $x = a$, $x = b$ and $y = 0$. If this point lands below $f(x)$, record the x coordinate, otherwise reject it. Thus the area for successfully picking points is $\int_a^b f(x)dx = 1$ which means the probability of doing so is

$$p = \frac{1}{\int_a^b M(x)dx}.$$

Furthermore, since this process of picking points is repeated until a success, the probability of such a success happening after k iterations is $(1 - p)^{k-1}p$ which is a geometric distribution. Finally, note that the closer $M(x)$ is to $f(x)$, the closer $\int_a^b M(x)dx$ is to 1, since $M(x) \geq f(x)$. This brings p closer and closer to 1 which decreases the likelihood of additional trials. □