The Two Hardest Topics on Exam P/1: Transformations and Order Statistics of Continuous Random Variables

From my experience in tutoring for exam P/1, it seems that these topics consistently are the most confusing topics for students. This document is meant to try to make these topics as clear as possible for students as they study for this exam.

I assume that the reader knows everything about expected values, probability density functions, (cumulative) distribution functions, and independence of continuous random variables in the exam P/1 syllabus. Although it is not very well emphasized in later exams, these two topics are very important in exams MLC, 3L, and C/4. Using the concepts presented here as it pertains to the material on these courses can help further understanding of the material and even speed up calculations (especially on exams MLC and 3L on multi-life models).

Assume all random variables in this document are continuous random variables.

1 Transformations

Let X be a random variable. A **transformation** of X is a function of X, which is also a random variable. We will also refer to this function of X as simply a "transformation." For example, $Y = 3X$, $Z = 3X + 4$, $W = |X|$ are all transformations of X and are also random variables. We are specifically interested in finding the probability density functions (PDFs) of transformations.

1.1 The Cumulative Distribution Function (CDF) Method

Since X is a continuous random variable, recall by definition that

$$f_X(x) = \frac{d}{dx}[F_x(x)] \tag{1}$$

and

$$P(X \leq x) = F_X(x). \tag{2}$$

Now if we consider $Y = g(X)$, Y is also a continuous random variable, and we can use (1) and (2) to find the PDF of Y, assuming we have the distribution function of X. This method of finding the density function of Y is known as the **CDF method**.

How to Do the CDF Method:

1. Find the domain of X and $F_X(x)$.
2. Find F_Y using the fact that $F_Y(y) = P(Y \leq y)$. Write $P(Y \leq y)$ in the form $P(X \leq h(y)) = F_X(h(y))$. Put in $h(y)$ for x. The resulting function is $F_Y(y)$.
3. Take the derivative of F_Y to find f_Y.
4. Find the domain of Y.

Example 1. Suppose X follows an exponential distribution with mean 2. Let $Y = 3X + 5$. Find the PDF of Y.

Solution. We follow the stepwise process for the CDF method.

1. $F_X(x) = 1 - e^{-x/2}$ for $x > 0$.[1]
2. Note that $P(Y \leq y) = P(3X + 5 \leq y) = P\left(X \leq \frac{y-5}{3}\right) = F_X\left(\frac{y-5}{3}\right) =$
$$1 - \exp\left[-\frac{\left(\frac{y-5}{3}\right)}{2}\right] = 1 - e^{-(y-5)/6} = F_Y(y).$$ [2]
3. The derivative of the above is $f_Y(y) = \frac{1}{6}e^{-(y-5)/6}$.
4. Since $X > 0$, we have $Y = 3X + 5 > 3 * 0 + 5 = 5$. So the domain of Y is $Y > 5$. $\square$

We can use the information about Y found above to calculate anything that we are interested in that has to do with Y: expected value, variance, probabilities in intervals, the CDF, moment-generating function (MGF), etc. Note that the CDF is the most general method presented in exam P/1 for finding the PDF of a transformation.

1.2 The Method of Transformations (Univariate)

Let X be a random variable and $Y = g(X)$ be strictly increasing/decreasing (or an invertible function) in the domain of X. Then

$$f_Y(y) = f_X(x(y)) \left| \frac{d}{dy}[x(y)] \right|, \tag{3}$$

where $x(y) = g^{-1}(y)$ is "x in terms of y."

Proof. Assume Y is strictly increasing. Since Y is invertible, we can write $X = g^{-1}(Y)$. So

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)) \\ &= F_X(x(y)) \text{ since } X = g^{-1}(Y). \end{aligned}$$

Now $f_Y(y) = \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}[F_X(x(y))] = f_X(x(y)) * \frac{d}{dy}[x(y)]$ by the chain rule.

[1] You should have this memorized for exam P/1.

[2] Note that $e^x = \exp(x)$.

If Y is strictly decreasing, $X = g^{-1}(Y)$ is decreasing and

$$\begin{aligned} F_Y(y) &= P(g(X) \leq y) \\ &= P(X \geq g^{-1}(y)) \\ &= 1 - P(X < g^{-1}(y)) \\ &= 1 - F_X(g^{-1}(y)). \end{aligned}$$

Thus, $f_Y(y) = \dfrac{d}{dy}[F_Y(y)] = \dfrac{d}{dy}[1 - F_X(g^{-1}(y))] = -f_X(g^{-1}(y))\dfrac{d}{dy}[g^{-1}(y)] = -f_X(x(y))\dfrac{d}{dy}[x(y)]$. Since $\dfrac{d}{dy}[x(y)]$ is negative (since X is decreasing), multiplying by the negative 1 will make it positive, and then, we have $f_Y(y) = f_X(x(y))\left|\dfrac{d}{dy}[x(y)]\right|$. □

How to Do the Method of Transformations:

1. Find the domain of X and the inverse of $Y = g(X)$. If Y is not invertible in the domain of X, then the method of transformations cannot be applied.
2. Find $f_X(x)$.
3. Apply the formula $f_Y(y) = f_X(x(y))\left|\dfrac{d}{dy}[x(y)]\right|$, where $x(y)$ is "x in terms of y."
4. Find the domain of Y.

Example 2. Apply the method of transformations to example 1.

Solution. The domain of X is $X > 0$. Now $Y = 3X + 5$ is invertible, since it is strictly increasing (with $\dfrac{d}{dX}[Y] = 3 > 0$ for all X).

Since $f_X(x) = \dfrac{1}{2}e^{-(1/2)x}$ for $x > 0$, it follows that since $Y = 3X + 5$, $X = \dfrac{Y-5}{3}$ and $f_Y(y) = f_X\left(\dfrac{y-5}{3}\right)\left|\dfrac{d}{dy}\left[\dfrac{y-5}{3}\right]\right| = \dfrac{1}{2}e^{-(1/2)(y-5)/3}\left|\dfrac{1}{3}\right| = \dfrac{1}{6}e^{-(y-5)/6}$. It was shown in the solution to example 1 that the domain of Y is $Y > 5$. □

Notice that examples 1 and 2 both provide the same answer.

1.3 Method of Jacobians

In order to understand this section, we will need to use some linear algebra. If we consider a two-by-two matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

the **determinant** of A, denoted $\det(A)$, is given by $ad - bc$.

Let X and Y be random variables, and $W_1 = W_1(X, Y)$, $W_2 = W_2(X, Y)$ be functions of X and Y. The two-by-two **Jacobian** matrix, denoted J, of W_1 and W_2 is

$$J = \begin{bmatrix} \frac{\partial}{\partial w_1}[x(w_1, w_2)] & \frac{\partial}{\partial w_2}[x(w_1, w_2)] \\ \frac{\partial}{\partial w_1}[y(w_1, w_2)] & \frac{\partial}{\partial w_2}[y(w_1, w_2)] \end{bmatrix}.$$

It can be shown that if W_1 and W_2 are one-to-one functions of X and Y,

$$f_{W_1,W_2}(w_1, w_2) = f_{X,Y}[x(w_1, w_2), y(w_1, w_2)]|\det(J)|. \tag{4}$$

How to Do the Method of Jacobians:

1. Find the domains of W_1 and W_2.
2. Check that W_1 and W_2 are one-to-one functions of X and Y.
3. Find $x(w_1, w_2)$ and $y(w_1, w_2)$.
4. Find
$$J = \begin{bmatrix} \frac{\partial}{\partial w_1}[x(w_1, w_2)] & \frac{\partial}{\partial w_2}[x(w_1, w_2)] \\ \frac{\partial}{\partial w_1}[y(w_1, w_2)] & \frac{\partial}{\partial w_2}[y(w_1, w_2)] \end{bmatrix}.$$
5. Find $\det(J)$.
6. Input into the formula
$f_{W_1,W_2}(w_1, w_2) = f_{X,Y}[x(w_1, w_2), y(w_1, w_2)]|\det(J)|$.

Example 3. Suppose X and Y are independent exponential random variables with means 1 and 2, respectively. That is, $f_{X,Y}(x,y) = \frac{1}{2}e^{-x}e^{-y/2}$. If $W = X+Y$ and $Z = 2Y$, find the joint density function of W and Z.

Solution.

1. Since $W = X+Y > 0+0 = 0$, $W > 0$. Also, $Z = 2Y > 2*0 = 0$, $Z > 0$.

2. You can check this step. (Both W and Z are one-to-one functions of X and Y.)

3. Notice that $Z = 2Y \implies Y = \frac{Z}{2}$. So $W = X+Y = X+\frac{Z}{2}$, which implies that $X = W-\frac{Z}{2}$. Hence $y(w,z) = \frac{z}{2}$ and $x(w,z) = w - \frac{z}{2}$.

4. $$J = \begin{bmatrix} \frac{\partial}{\partial w}[x(w,z)] & \frac{\partial}{\partial z}[x(w,z)] \\ \frac{\partial}{\partial w}[y(w,z)] & \frac{\partial}{\partial z}[y(w,z)] \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}.$$

5. $\det(J) = 1 * \frac{1}{2} - \frac{-1}{2}(0) = \frac{1}{2}$. Of course, $\left|\frac{1}{2}\right| = \frac{1}{2}$.

6. $f_{W,Z}(w,z) = f_{X,Y}\left(w - \frac{z}{2}, \frac{z}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{2}e^{-\left(w-\frac{z}{2}\right)}e^{-(z/2)/2} = \frac{1}{2}e^{-w}e^{z/2}e^{-z/4} = \frac{1}{2}e^{-w}e^{z/4} = \frac{1}{2}e^{-w+(z/4)}$. $\square$

2 Order Statistics

Let $X = \{X_1, X_2, \cdots, X_n\}$ be a set of independent random variables. We are interested in the order statistics $X_{(1)} = \min[X]$ and $X_{(n)} = \max[X]$. (If you would like information on other order statistics, consult a probability textbook.)

It is our task to find the PDF and CDF of both $X_{(1)}$ and $X_{(n)}$. Note that

$$F_{X_{(1)}}(x) = P(\min[X] \leq x)$$

gives us no information (by itself) whatsoever about all of the random variables in the set X. But, if we looked at

$$P(\min[X] > x),$$

this <u>does</u> give us information about all of the random variables in the set X. If the minimum of X is greater than x, what does that tell us? Every random variable in X must also be greater than x.

So, in finding the CDF of $X_{(1)}$, we write

$$\begin{aligned} F_{X_{(1)}}(x) &= P(\min[X] \leq x) \\ &= 1 - P(\max[X] > x) \\ &= 1 - P(\text{all of } X_1, X_2, \cdots, X_n > x) \\ &= 1 - P(X_1 > x)P(X_2 > x)\cdots P(X_n > x) \text{ by independence.} \end{aligned} \tag{5}$$

If all random variables in X are identically distributed with CDF F, then $P(X_i > x) = 1 - F(x)$ for all i and

$$F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n. \tag{6}$$

Take the derivatives of what you get from either (5) or (6) above to get $f_{X_{(1)}}(x)$. Be sure to find the appropriate domain for $X_{(1)}$.

My suggestion: do not memorize equation (6). As not each random variable in X will always be identically distributed, (6) will not work every time for an exam question. I passed P/1 by memorizing the methods to get (5).

Example 4. Let X_1, X_2, X_3 be exponentially distributed with means 1, 2, and 3 respectively. Find the PDF of $X_{(1)} = \min\{X_1, X_2, X_3\}$.

Solution.

$$\begin{aligned} F_{X_{(1)}}(x) &= P(\min[X] \leq x) \\ &= 1 - P(X_1 > x)P(X_2 > x)P(X_3 > x) \text{ by (5)} \\ &= 1 - e^{-x}e^{-(1/2)x}e^{-(1/3)x} \\ &= 1 - e^{-(11/6)x}. \end{aligned}$$

Now $\dfrac{d}{dx}[F_{X_{(1)}}(x)] = \dfrac{11}{6}e^{-(11/6)x} = f_{X_{(1)}}(x)$ for $x > 0$ (the minimum of X_1, X_2, X_3 would have domain > 0 since all three random variables have domain > 0). □

Now we find the CDF of $X_{(n)}$. If the maximum of X is less than or equal to some value x, then every value in X is less than x. So

$$\begin{aligned} F_{X_{(n)}}(x) &= P(\max[X] \leq x) \\ &= P(\text{all of } X_1, X_2, \cdots, X_n \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x)\cdots P(X_n \leq x) \text{ by independence.} \end{aligned} \tag{7}$$

If all random variables are identically distributed with CDF F, then $P(X_i \leq x) = F(x)$ and

$$F_{X_{(n)}}(x) = [F(x)]^n. \tag{8}$$

Take the derivatives of what you get from either (7) or (8) to get $f_{X_{(n)}}(x)$. Be sure to find the appropriate domain for $X_{(n)}$. Similarly, I would suggest not memorizing (8), but the methods to get to (7).

Example 5. Using the information in example 4, find the CDF of $X_{(n)} = \max\{X_1, X_2, X_3\}$.

Solution.

$$\begin{aligned}
F_{X_{(n)}}(x) &= P(\max[X] \leq x) \\
&= P(X_1 < x)P(X_2 < x)P(X_3 < x) \text{ by (7)} \\
&= [1 - e^{-x}][1 - e^{-(1/2)x}][1 - e^{-(1/3)x}] \text{ for } x > 0. \ \square
\end{aligned}$$

Order Statistics:

1. Find the domain of the order statistic.
2. Find the CDF of the order statistic you need by using the methods to get (5) or (7).
3. Take the derivative of this function to get the PDF of the order statistic.