Stat 134: Change of Variables and Operations - Review

Conceptual Review

a. Let X be a discrete random variable and set $Y = g(X)$, what is a formula for $\mathbb{P}(Y = y)$?

Solution: $\mathbb{P}(Y = y) = \sum_{x:g(x)=y} \mathbb{P}(X = x)$

b. Let now X be a continuous random variable with density f_X and set again $Y = g(X)$. What is a formula for the density f_Y of Y?

Solution: $f_Y(y) = \sum_{\{x:g(x)=y\}} \frac{f_X(x)}{|g'(x)|}$, where the derivative of g is taken with respect to x. While the formula on the left looks as if it had the variable x, this is not the case because we express the x we sum over in terms of y.

c. Which steps do we need to follow when applying this formula?

Solution:

Step 1: Determine the range of Y.

Step 2: Find the set $\{x : g(x) = y\}$ (This means find all points that map to y under g).

Step 3: Compute the derivative of g.

Step 4: Plug these into the change of variable formula, being careful about the support of f_X.

Step 5: If you have time check if the density you found integrates to 1.

d. Is it necessary to do a change of variables in order to compute $\mathbb{E}[g(X)]$?

Solution: No, in this case we use that for continuous random variables, where $f_X(x)$ is the density of X, $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$ and for discrete random variables $\mathbb{E}[g(X)] = \sum_{x \in \text{ range of } X} g(x)\mathbb{P}(X = x)$.

e. What is the density of a sum of two continuous random variables $X + Y$?

Solution: If (X, Y) has the density $f(x, y)$ for $(x, y) \in \mathbb{R}^2$,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z - x) dx = \int_{-\infty}^{\infty} f(y - z, y) dy.$$

f. If X and Y are discrete, how can we find an expression for $\mathbb{P}(X+Y=z)$?

Solution:

$$\mathbb{P}(X+Y=z)=\sum_{x\in \text{ range of X}}\mathbb{P}(X=x,Y=z-x)=\sum_{y\in \text{ range of Y}}\mathbb{P}(X=z-y,Y=y)$$

g. What is the density of the ratio of two positive continuous random variables $\frac{X}{Y}$?

Solution: If (X,Y) has the density $f_{X,Y}(x,y)$ for $x,y>0$,

$$f_{\frac{X}{Y}}(z)=\int_0^\infty yf(yz,y)dy.$$

Problem 1

Let X and Y be exponentially distributed with parameters λ, resp. μ. Find the density of $R = \frac{X}{Y}$.

1. Solve this question using the formula for densities of ratios.

   Solution: Using our formula we get for $r > 0$,

$$\begin{aligned} f_R(r) &= \int_0^\infty y\lambda e^{-\lambda yr}\mu e^{-\mu y} dy \\ &= \lambda\mu \int_0^\infty ye^{-(\lambda r+\mu)y} dy \\ &= \frac{\lambda\mu}{\lambda r + \mu}\int_0^\infty y(\lambda r + \mu)e^{-(\lambda r+\mu)y} dy \\ &= \frac{\lambda\mu}{(\lambda r + \mu)^2}. \end{aligned}$$

2. Try to relate the problem to competing exponentials.

   Solution: Alternatively we can note that $\mathbb{P}\left(\frac{X}{Y} < r\right) = \mathbb{P}(X < rY)$. Now rY is exponential with rate $\frac{\mu}{r}$ (you can prove this using a change of variables formula), so the above probability is one about competing exponentials. We know that $\mathbb{P}(X < rY) = \frac{\lambda}{\lambda+\frac{\mu}{r}} = \frac{r\lambda}{r\lambda+\mu}$.

   So $f_R(r) = \frac{d}{dr}F_R(r) = \frac{d}{dr}\frac{r\lambda}{r\lambda+\mu} = \frac{\lambda\mu}{(r\lambda+\mu)^2}$.

3. Find the density by first computing the cdf.

   Solution: The cdf is equal to the integral of the joint density over the region where $X/Y < r$, which is equivalent to $X < rY$. Since exponentials are positive, the lower bounds of integration are 0.

$$
\begin{aligned}
F_R(r) &= \int_0^\infty \int_0^{ry} f_{X,Y}(x,y)dxdy \\
&= \int_0^\infty \int_0^{ry} \lambda e^{-\lambda x} \mu e^{-\mu y} dxdy \\
&= \int_0^\infty \mu e^{-\mu y}(1 - e^{-\lambda r y})dy \\
&= 1 - \mu \int_0^\infty e^{-(\lambda r + \mu)y} dy \\
&= 1 - \frac{\mu}{\lambda r + \mu} \\
&= \frac{\lambda r}{\lambda r + \mu}
\end{aligned}
$$

Problem 2

Assume that we first flip a coin until we get heads, where the probability of getting head at a toss is p. Let T be the number of tosses we need. Given $T = t$, we toss a coin with success probability $\frac{1}{t}$ until we get heads for the first time. Let S denote the number of tosses we need this time. What is the distribution of $Z = T + S$?

Step 1: What is the range of Z?

> **Solution:** Since $T \in \{1, 2, \dots\}$ and $S \in \{1, 2, \dots, t\}$ the range of Z is $\{2, 3, \dots\}$.

Step 2: For z in the range of Z, find an expression for $\mathbb{P}(Z = z)$.

> **Solution:** For $z \in \{1, 2, \dots\}$,
>
> $$\mathbb{P}(Z = z) = \sum_{t=1}^{\infty} \mathbb{P}(T = t, S = z - t).$$
>
> $\mathbb{P}(T = t, X = z - t) > 0$ iff $\begin{cases} 1 \le t \\ 1 \le z - t \end{cases}$ iff $\begin{cases} 1 \le t \\ t \le z - 1 \end{cases}$ iff $1 \le t \le z - 1$.
>
> For all t that satisfy the above
>
> $$\mathbb{P}(T = t, S = z - t) = (1 - p)^{t-1} p \left(1 - \frac{1}{t}\right)^{z-t-1} \frac{1}{t},$$
>
> so
>
> $$\mathbb{P}(Z = z) = \sum_{t=1}^{z-1} (1 - p)^{t-1} p \left(1 - \frac{1}{t}\right)^{z-t-1} \frac{1}{t}.$$

Problem 3

Let X and Y be i.i.d. uniform on $\left(0, e^{-1}\right)$. Determine the distribution of $\log(XY)$.

Step 1: This is not an operation of two random variables we immediately know how to deal with. Try to get it into a different form.

Solution: We have that $\log(XY) = \log(X) + \log(Y)$. This is a sum of two independent random variables, so if we can find the densities of $\log(X)$ and $\log(Y)$ we can use our formula for sums of random variables.

Step 2: Find the density of $V = \log(X)$.

Solution:

1. The range of X is $\left(0, e^{-1}\right)$, so the range of V is $(-\infty, -1)$.
2. $\log(x) = v \Leftrightarrow x = e^{v}$, so $\{x : g(x) = v\} = \{e^{v}\}$
3. $g'(x) = \frac{1}{x}$ for $x > 0$.
4. Using these and the change of variables formula we get for $v < -1$,

$$f_V(v) = \frac{f_X(e^{v})}{\frac{1}{e^{v}}} = \frac{1}{e^{-1}} e^{v} = e^{v+1}$$

and $f_V(v) = 0$ otherwise.

Since X and Y are identically distributed, $W = \log(Y)$ has the same density.

Step 3: Can you recognize the distribution of V? If yes, use this to determine the distribution of $Z = V + W$. If not, skip to the next step.

Solution: The density of V is kind of resembles that of an exponential r.v., this suggests that we might be able to express V in terms of an exponential random variable. Actually the random variable $-V - 1$ is exponentially distributed: For $v > 0$,

$$\mathbb{P}(-V - 1 > v) = \mathbb{P}(V + 1 < -v) = \mathbb{P}(V < -1 - v)$$
$$= \int_{-\infty}^{-1-v} e^{y+1} dy = e^{y+1}\Big|_{-\infty}^{-1-v} = e^{-v},$$

which is the survival function of an exponential random variable of rate 1.

This implies that the density of $-V-1+(-W-1) = T$ is $gamma(2,1)$. Since $V+W = -T-2$, we can compute the density of $V+W$ using a change of variables formula:

$$f_{V+W}(z) = f_{-T-2}(z) = f_T(-z-2)$$

The density of a gamma random variable is positive on $(0,\infty)$, so for $z < -2$, it holds that

$$f_T(-z-2) = -(z+2)e^{-(-z-2)} = -(z+2)e^{z+2}.$$

So $f_{V+W}(z) = \begin{cases} -(z+2)e^{z+2} & \text{for } z < -2 \\ 0 & \text{otherwise.} \end{cases}$

Step 3' Use the formula for densities of sums of random variables to find the density of $Z = V+W$.

Solution:

1. Find the range of Z: since the ranges of V and W are $(-\infty,-1)$, $-\infty < Z < -2$.
2. For values in the range of Z, find the density of Z: For $z < -2$, since V and W are independent,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{V,W}(v, z-v)dv = \int_{-\infty}^{\infty} f_V(v)f_W(z-v)dv.$$

We now need to determine the bounds of integration. The densities must both be non-zero for the product to be non-zero. This holds if

$$\begin{cases} v < -1 \\ z-v < -1 \end{cases} \Leftrightarrow \begin{cases} v < -1 \\ z+1 < v \end{cases} \Leftrightarrow z+1 < v < -1.$$

Since for $z < -2$, it holds that $z+1 < -1$, this is a valid interval. Thus for $z < -2$

$$f_Z(z) = \int_{z+1}^{-1} e^{v+1}e^{z-v+1}dv = e^{z+2}v|_{z+1}^{-1} = -(z+2)e^{z+2}.$$

Otherwise the density is 0.