STAT 30400 : Distribution Theory

Fall 2019

•

Homework 3

•

Solutions by

Jinhong Du

12243476

1. (10 pts) Let $X_1 \sim \text{Gamma}(r_1, \lambda)$ and $X_2 \sim \text{Gamma}(r_2, \lambda)$ be independent random variables, and let $Y = X_1 + X_2$ and $Z = \frac{X_1}{X_1+X_2}$.

   (a) Find the joint density of Y and Z, and find the marginal densities of Y and Z. Identify the distributions of Y and Z. Note that the $\text{Gamma}(r, \lambda)$ distribution has density function,

$$f(x) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1}.$$

The inverse transformation from (X_1, X_2) to (Y_Z) is given by

$$\begin{cases} X_1 = YZ \\ X_2 = Y(1-Z) \end{cases}.$$

The determinant of Jacobian of this inverse transformation is

$$J = \left| \begin{pmatrix} Z & Y \\ 1-Z & -Y \end{pmatrix} \right| = -YZ - Y(1-Z) = -Y.$$

So the joint density of Y and Z is

$$\begin{aligned} f_{(Y,Z)}(y,z) &= f_{X_1}(yz) f_{X_2}(y(1-z)) | - y| \mathbb{1}_{\{y>0, yz>0, y(1-z)>0\}} \\ &= \frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)} e^{-\lambda y} (yz)^{r_1-1} [y(1-z)]^{r_2-1} \mathbb{1}_{\{y>0, 0<z<1\}} \\ &= \frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)} e^{-\lambda y} y^{r_1+r_2-1} z^{r_1-1} (1-z)^{r_2-1} \mathbb{1}_{\{y>0, 0<z<1\}}. \end{aligned}$$

Integrating on y,

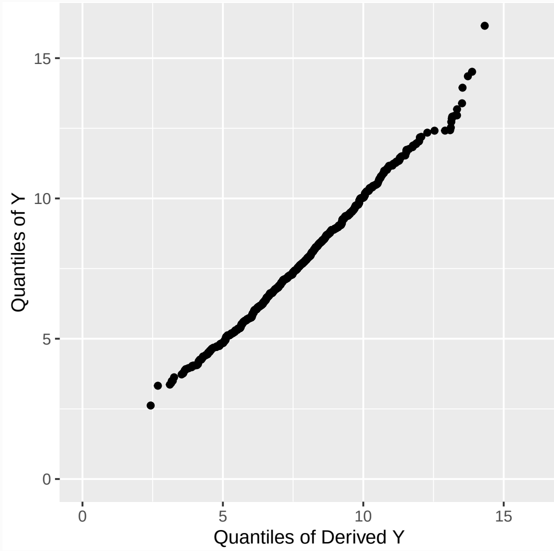
$$\begin{aligned} f_Z(z) &= \int_0^\infty \frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)} e^{-\lambda y} y^{r_1+r_2-1} z^{r_1-1} (1-z)^{r_2-1} \mathbb{1}_{\{0<z<1\}} \mathrm{d}y \\ &= \frac{\Gamma(r_1+r_2)}{\Gamma(r_1)\Gamma(r_2)} z^{r_1-1} (1-z)^{r_2-1} \mathbb{1}_{\{0<z<1\}} \int_0^\infty \frac{\lambda^{r_1+r_2}}{\Gamma(r_1+r_2)} e^{-\lambda y} y^{r_1+r_2-1} \mathrm{d}y \\ &= \frac{\Gamma(r_1+r_2)}{\Gamma(r_1)\Gamma(r_2)} z^{r_1-1} (1-z)^{r_2-1} \mathbb{1}_{\{0<z<1\}} \end{aligned}$$

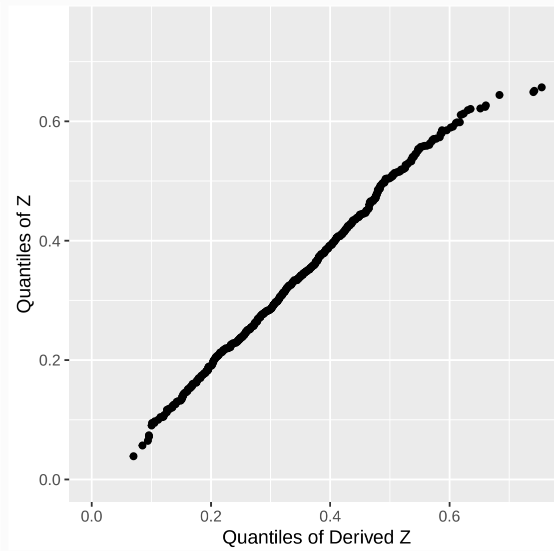
So

$$f_Y(y) = \frac{\lambda^{r_1+r_2}}{\Gamma(r_1+r_2)} e^{-\lambda y} y^{r_1+r_2-1} \mathbb{1}_{\{y>0\}}$$

with $f_{(Y,Z)}(y,z) = f_Y(y) f_Z(z)$, i.e., Y and Z are independent. $Y \sim \Gamma(r_1+r_2, \lambda)$ and $Z \sim \text{Beta}(r_1, r_2)$.

(b) Use QQ plots on simulated data to demonstrate that your marginal densities are correct. Show the plots and the work you have done to construct them.





Code:

```
set.seed(1)
n <- 1000
r_1 <- 5
r_2 <- 10
lambda <- 2
X_1 <- rgamma(n, r_1, lambda)
X_2 <- rgamma(n, r_2, lambda)
Y <- X_1 + X_2
Z <- X_1 / Y

Y_derived <- rgamma(n, r_1+r_2, lambda)
Z_derived <- rbeta(n, r_1, r_2)

library(ggplot2)
df <- as.data.frame(qqplot(Y_derived, Y, plot.it=FALSE));
ggplot(df) + geom_point(aes(x=x, y=y)) +
    coord_fixed(ratio = 1,
                xlim=c(0,max(Y,Y_derived)), ylim=c(0,max(Y,Y_derived))) +
    xlab('Quantiles of Derived Y') +
    ylab('Quantiles of Y')
df <- as.data.frame(qqplot(Z_derived, Z, plot.it=FALSE));
ggplot(df) + geom_point(aes(x=x, y=y)) +
    coord_fixed(ratio = 1,
                xlim=c(0,max(Z,Z_derived)), ylim=c(0,max(Z,Z_derived))) +
    xlab('Quantiles of Derived Z') +
    ylab('Quantiles of Z')
```

2. (10 pts) Random variables $X_1, \dots, X_n$ are said to be *exchangeable* if the joint distribution of $(X_1, \dots, X_n)$ is the same as the distribution of $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$, for any permutation σ of $\{1, \dots, n\}$.

(a) Give an example of two random variables X_1 and X_2 such that they have discrete distributions, are exchangeable and are not independent.

Let $X_1 \sim Bernoulli(\frac{1}{2})$ and $X_2 = 1 - X_1$. Obviously, X_1 and X_2 are not independent. Then the joint probability mass function is given by

$$p_{X_1,X_2}(x_1, x_2) = \begin{cases} \frac{1}{2} & , (x_1, x_2) = (1, 0) \\ \frac{1}{2} & , (x_1, x_2) = (0, 1) \\ 0 & , otherwise \end{cases}$$

Then we have $p_{X_1,X_2}(x_1, x_2) = p_{X_2,X_1}(x_2, x_1)$, i.e., X_1 and X_2 are exchangeable.

(b) Let $Y_1, Y_2, \dots, Y_{n+1}$ denote iid random variables, and define

$$X_j = Y_j Y_{j+1}, \qquad j = 1, \dots, n.$$

Do $X_1, \dots, X_n$ have the same (marginal) distribution? Are $X_1, \dots, X_n$ iid? Are $X_1, \dots, X_n$ exchangeable?

Since $Y_1, Y_2, \dots, Y_{n+1}$ are iid random variables, $Y_j Y_{j+1}$ $(j = 1, \dots, n)$ have same distribution. So $X_1, \dots, X_n$ have the same (marginal) distribution.
Let $Y_1, \dots, Y_4 \overset{iid}{\sim} Bernoulli(1, \frac{1}{3})$, then $X_1 = Y_1 Y_2$ and $X_2 = Y_2 Y_3$ take values in $\{0, 1\}$ and

$$\begin{aligned} \mathbb{P}(X_i = 1) &= \mathbb{P}(Y_i Y_{i+1} = 1) \\ &= \mathbb{P}(Y_i = 1, Y_{i+1} = 1) \\ &= \mathbb{P}(Y_i = 1)\mathbb{P}(Y_{i+1} = 1) \\ &= \frac{1}{9}, \\ \mathbb{P}(X_1 = 1, X_2 = 1) &= \mathbb{P}(Y_1 Y_2 = 1, Y_2 Y_3 = 1) \\ &= \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = 1) \\ &= \mathbb{P}(Y_1 = 1)\mathbb{P}(Y_2 = 1)\mathbb{P}(Y_3 = 1) \\ &= \frac{1}{27}, \end{aligned}$$

so

$$\mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) \neq \mathbb{P}(X_1 = 1, X_2 = 1),$$

i.e., X_1 and X_2 are not independent.

Solution (cont.)

$$\begin{aligned}
\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 0) &= \mathbb{P}(Y_1Y_2 = 1, Y_2Y_3 = 1, Y_3Y_4 = 0) \\
&= \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = 1, Y_4 = 0) \\
&= \mathbb{P}(Y_1 = 1)\mathbb{P}(Y_2 = 1)\mathbb{P}(Y_3 = 1)\mathbb{P}(Y_4 = 0) \\
&= \frac{2}{3^4} \\
\mathbb{P}(X_1 = 1, X_3 = 1, X_2 = 0) &= \mathbb{P}(Y_1Y_2 = 1, Y_2Y_3 = 0, Y_3Y_4 = 1) \\
&= 0
\end{aligned}$$

Since $\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 0) \neq \mathbb{P}(X_1 = 1, X_3 = 1, X_2 = 0)$, $X_1, \ldots, X_3$ are not exchangeable.

3. (10 pts) Let's denote with $Q_{n,m}$ the quantile function of the $F_{n,m}$ distribution.

(a) Show that

$$Q_{n,m}(\alpha) = \frac{1}{Q_{m,n}(1-\alpha)}, \qquad \forall \alpha \in (0,1).$$

Proof. Let $f_{n,m}(x)$ and $F_{n,m}(x)$ be the density function and cumulative distribution function of $F_{n,m}$ distribution respectively, then

$$\begin{aligned}
f_{n,m}(x) &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} \left(1+\frac{n}{m}x\right)^{-\frac{m+n}{2}} \mathbb{1}_{\{x\geq 0\}} \\
F_{n,m}(x) &= \int_0^x \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} t^{\frac{n}{2}-1} \left(1+\frac{n}{m}t\right)^{-\frac{m+n}{2}} \mathrm{d}t \\
&= \int_0^{\frac{n}{m}x} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} t^{\frac{n}{2}-1} (1+t)^{-\frac{m+n}{2}} \mathrm{d}t \\
&= 1 - \int_{\frac{n}{m}x}^{+\infty} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} t^{\frac{n}{2}-1} (1+t)^{-\frac{m+n}{2}} \mathrm{d}t \\
&\overset{w=\frac{1}{t}}{=\!=\!=} 1 + \int_{\frac{m}{nx}}^{0} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} w^{\frac{m}{2}-1} (1+w)^{-\frac{m+n}{2}} \mathrm{d}w \\
&= 1 - \int_0^{\frac{1}{x}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} w^{\frac{m}{2}-1} \left(1+\frac{m}{n}w\right)^{-\frac{m+n}{2}} \mathrm{d}w \\
&= 1 - F_{m,n}\left(\frac{1}{x}\right), \qquad x > 0
\end{aligned}$$

So,

$$\begin{aligned}
Q_{n,m}(\alpha) &= \inf_{x>0}\{F_{n,m}(x) \geq \alpha\} \\
&= \inf_{x>0}\left\{1 - F_{m,n}\left(\frac{1}{x}\right) \geq \alpha\right\} \\
&= \inf_{x>0}\left\{F_{m,n}\left(\frac{1}{x}\right) \leq 1-\alpha\right\} \\
&= \frac{1}{\sup\limits_{x>0}\{F_{m,n}(x) \leq 1-\alpha\}} \\
&= \frac{1}{\inf\limits_{x>0}\{F_{m,n}(x) \geq 1-\alpha\}} \\
&= \frac{1}{Q_{m,n}(1-\alpha)}
\end{aligned}$$

where the second to last equality holds since $F_{m,n}(x)$ is absolutely continuous. $\square$

(b) Find the (analytical) relation between the quantile function of the t_n distribution and $Q_{1,n}$. Also, find the relation between the quantile function of the t_n distribution and $Q_{n,1}$.

Let $f_{t_n}(x)$ and $F_{t_n}(x)$ be the density function and cumulative distribution function of t_n distribution respectively.

$$\begin{aligned}
f_{t_n}(x) &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \\
f_{1,n}(x) &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{n}\right)^{\frac{1}{2}} x^{\frac{1}{2}-1}\left(1+\frac{1}{n}x\right)^{-\frac{n+1}{2}} \mathbb{1}_{\{x\geq 0\}} \\
F_{1,n}(x) &= \int_0^x \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{n}\right)^{\frac{1}{2}} t^{\frac{1}{2}-1}\left(1+\frac{1}{n}t\right)^{-\frac{n+1}{2}} \mathrm{d}t \\
&\overset{u=\sqrt{t}}{=\!=} 2\int_0^{\sqrt{x}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{u^2}{n}\right)^{-\frac{n+1}{2}} \mathrm{d}u \\
&= F_{t_n}(\sqrt{x}) - F_{t_n}(-\sqrt{x}) \\
&= 2F_{t_n}(\sqrt{x}) - 1, \qquad x > 0
\end{aligned}$$

So

$$\begin{aligned}
Q_{1,n}(\alpha) &= \inf_{x>0}\{F_{1,n}(x) \geq \alpha\} \\
&= \inf_{x>0}\left\{2F_{t_n}(\sqrt{x}) - 1 \geq \alpha\right\} \\
&= \inf_{x>0}\left\{F_{t_n}(\sqrt{x}) \geq \frac{\alpha+1}{2}\right\} \\
&= Q_{t_n}\left(\frac{\alpha+1}{2}\right)^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
Q_{n,1}(\alpha) &= \frac{1}{Q_{1,n}(1-\alpha)} \\
&= \frac{1}{Q_{t_n}\left(1-\frac{\alpha}{2}\right)^2}
\end{aligned}$$

4. (10 pts) Let X be an integrable real random variable with distribution function F, quantile function Q and mean μ. The quantity $\mathbb{E}|X-\mu|$ is called the mean absolute deviation of X from the mean. Prove that,

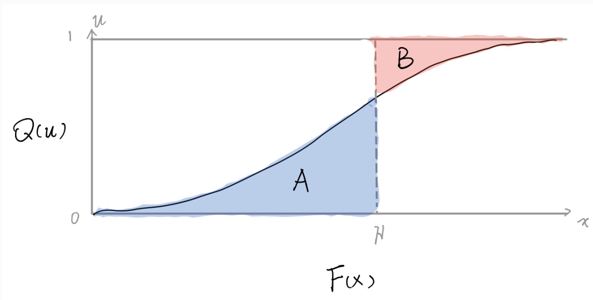
$$\mathbb{E}|X-\mu| = \int_0^1 |Q(u)-\mu|\mathrm{d}u = 2\int_{-\infty}^{\mu} F(x)\mathrm{d}x = 2\int_{\mu}^{\infty}(1-F(x))\mathrm{d}x.$$

Proof. Since X is integrable, $\mathbb{E}|X-\mu| \le \mathbb{E}|X| + \mu < \infty$, i.e. $|X-\mu|$ is also integrable. Since $X = F^{-}(U) = Q(U)$ where $U \sim Uniform(0,1)$, we have

$$\mathbb{E}|X-\mu| = \mathbb{E}|Q(U)-\mu| = \int_0^1 |Q(u)-\mu|\mathrm{d}u$$

Let $u_0 = F(\mu)$,

$$\int_0^1 |Q(u)-\mu|\mathrm{d}u = \int_0^{u_0}(\mu - Q(u))\mathrm{d}u + \int_{u_0}^1 (Q(u)-\mu)\mathrm{d}u = |A| + |B| < \infty$$



i.e., the integral $\int_0^{u_0}(\mu - Q(u))\mathrm{d}u$ and $\int_{u_0}^1 (Q(u)-\mu)\mathrm{d}u$ are the areas of the region

$$A = \{(x,u) : 0 \le u \le F(\mu), -\infty \le x \le \mu\}, \qquad B = \{(x,u) : F(\mu) \le u \le 1, \mu \le x \le \infty\},$$

respectively. By slicing A and B into infinitesimal vertical strips instead of horizontal ones, we can also compute $|A|$ and $|B|$ as

$$|A| = \int_{-\infty}^{\mu} F(x)\mathrm{d}x, \qquad |B| = \int_{\mu}^{\infty}[1-F(x)]\mathrm{d}x.$$

and

$$\begin{aligned}
|B| &= x[1-F(x)]\Big|_{\mu}^{\infty} - \int_{\mu}^{\infty} x\mathrm{d}[1-F(x)] \\
&= -\mu[1-F(\mu)] - \int_{\mu}^{\infty} x\mathrm{d}1 + \int_{\mu}^{\infty} x\mathrm{d}F(x) \\
&= \mu F(\mu) - \int_{-\infty}^{\infty} x\mathrm{d}F(x) + \int_{\mu}^{\infty} x\mathrm{d}F(x) \\
&= \mu F(\mu) - \int_{-\infty}^{\mu} x\mathrm{d}F(x) \\
&= \int_{-\infty}^{\mu} F(x)\mathrm{d}x = |A|
\end{aligned}$$

where $\int_{\mu}^{\infty} x\mathrm{d}1 = 0$ comes from the definition of Riemann-Stieltjes integrals. So

$$\begin{aligned}
\int_0^1 |Q(u)-\mu|\mathrm{d}u = |A| + |B| &= 2\int_{-\infty}^{\mu} F(x)\mathrm{d}x \\
&= 2\int_{\mu}^{\infty}(1-F(x))\mathrm{d}x.
\end{aligned}$$

$\square$

5. (10 pts) Let U be a random variable uniformly distributed on $(0,1)$, and let $X_\alpha = U^\alpha$, for $\alpha \in \mathbb{R}$. Find the density and distribution function of X_α. Use R to plot the density and distribution function of X_α for $\alpha = -1, \frac{1}{2}, 2$.

Since $\alpha \in \mathbb{R}$, let $\alpha = \frac{m}{n}$ where $m, n \in \mathbb{Z}$. The density function of U is $f_U(u) = \mathbb{1}_{(0,1)}$.

(1) If $\alpha > 0$, $X_\alpha = U^\alpha$ taking values from $(0,1)$,

$$U = (X_\alpha)^{\frac{1}{\alpha}}$$

Therefore,

$$f_{X_\alpha}(x) = f_U(x^{\frac{1}{\alpha}}) \left| \frac{1}{\alpha} x^{\frac{1}{\alpha}-1} \right| \mathbb{1}_{(0,1)} = \frac{1}{\alpha} x^{\frac{1}{\alpha}-1} \mathbb{1}_{(0,1)}$$

$$F_{X_\alpha}(x) = \begin{cases} 0 & , x \le 0 \\ \int_{-\infty}^{x} \frac{1}{\alpha} t^{\frac{1}{\alpha}-1} \mathbb{1}_{(0,1)} \mathrm{d}t = x^{\frac{1}{\alpha}} & , x \in (0,1) \\ 1 & , x \ge 1 \end{cases}$$

(2) If $\alpha < 0$, $X_\alpha = U^\alpha$ taking values from $(1,\infty)$,

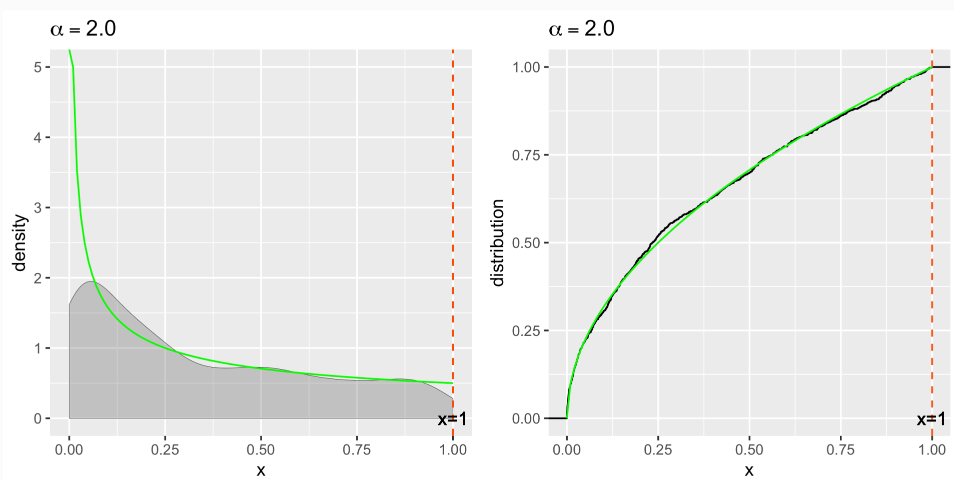
$$U = (X_\alpha)^{\frac{1}{\alpha}}$$

Therefore,

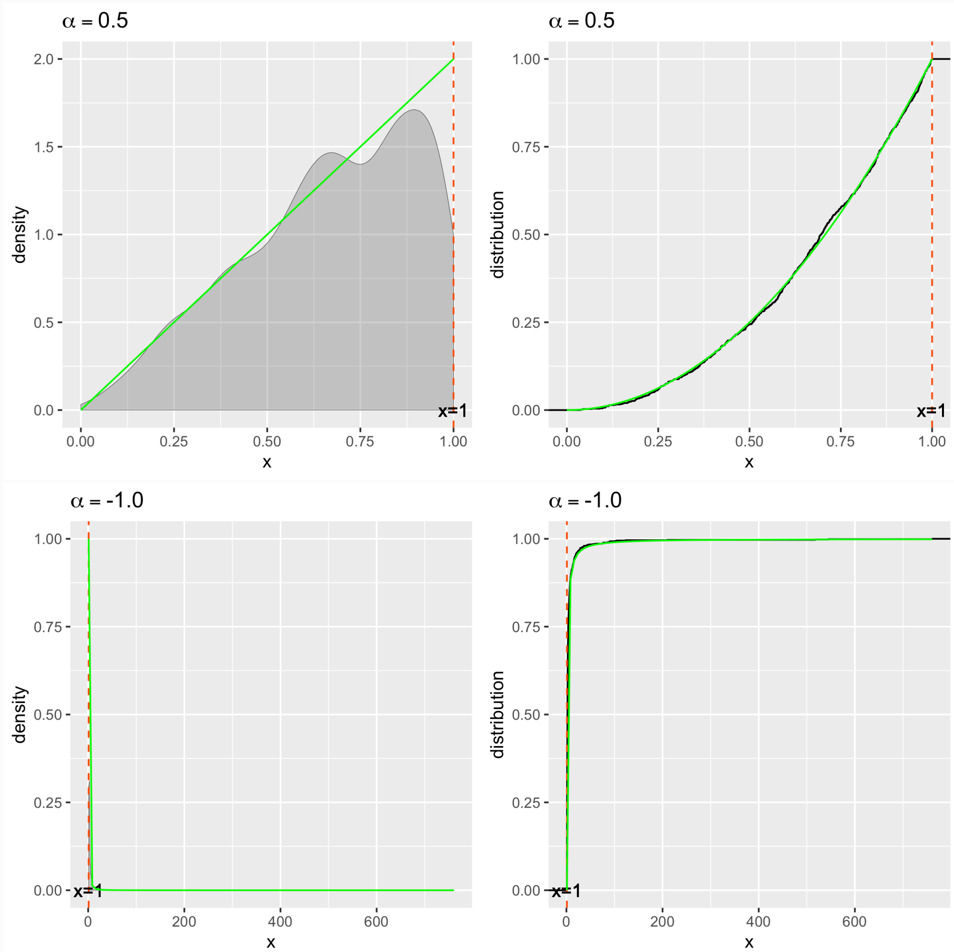
$$f_{X_\alpha}(x) = f_U(x^{\frac{1}{\alpha}}) \left| \frac{1}{\alpha} x^{\frac{1}{\alpha}-1} \right| \mathbb{1}_{(1,\infty)} = \frac{1}{-\alpha} x^{\frac{1}{\alpha}-1} \mathbb{1}_{(1,\infty)}$$

$$F_{X_\alpha}(x) = \begin{cases} 0 & , x \le 0 \\ \int_{-\infty}^{x} \frac{1}{-\alpha} t^{\frac{1}{\alpha}-1} \mathbb{1}_{(1,\infty)} \mathrm{d}t = 1 - x^{\frac{1}{\alpha}} & , x \in (1,\infty) \end{cases}$$

(3) If $\alpha = 0$, then $X_\alpha = 1$ has a singular dsitribution $F_{X_0}(x) = \mathbb{1}_{\{x \ge 1\}}$.



Solution (cont.)



The green lines denote the ground truth curve.

Code:

```
library(ggplot2)
library(latex2exp)
library(gridExtra)
make_plot <- function(alpha){
    set.seed(0)
    n <- 1000
    U <- runif(n)
    X <- U**alpha
    df <- data.frame(x=X)

    if (alpha<0){
        x_range <- c(1,max(X))
        distribution_function <- function(x)1-x^(1/alpha)
    }
```

Solution (cont.)

```
    else{
        x_range <- c(0,1)
        distribution_function <- function(x)x^(1/alpha)
    }
    density_function <- function(x)x^(1/alpha-1)/abs(alpha)

    p1 <- ggplot(df, aes(x=x)) +
        geom_density(fill='#868686FF', size=0.1, alpha = 0.4) +
        ggtitle(TeX(sprintf("$\\alpha = %2.1f$", alpha))) +
        geom_vline(xintercept = 1, linetype='dashed', color='#FC4E07') +
        geom_text(aes(x=1, label="x=1", y=0.0)) +
        ylab('density') + xlim(x_range)
    p1 <- p1 + stat_function(data=data.frame(x=x_range),
                             aes(x), fun=density_function, color='green')

    p2 <- ggplot(df, aes(x=x)) +
        stat_ecdf(geom = "step") +
        ggtitle(TeX(sprintf("$\\alpha = %2.1f$", alpha))) +
        geom_vline(xintercept = 1, linetype='dashed', color='#FC4E07') +
        geom_text(aes(x=1, label="x=1", y=0.0)) +
        ylab('distribution') + xlim(x_range)
    p2 <- p2 + stat_function(data=data.frame(x=x_range),
                             aes(x), fun=distribution_function, color='green')
    p <- grid.arrange(p1, p2, nrow = 1, widths=c(1,1))
    ggsave(filename = sprintf("result%0.1f.png", alpha), plot = p, width = 8, height = 4)
}

make_plot(-1)
make_plot(1/2)
make_plot(2)
```