The χ^2 distribution

The χ^2 distribution is important in fields such as econometrics and life modelling. Parameterised by the degrees of freedom, k, the distribution function is

$$F(x) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(k/2)},$$

where γ is the lower incomplete gamma function, and probability density function

$$f(x|k) = \frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}.$$

The χ^2 distribution is a special case of the gamma distribution with rate $\theta = \frac{1}{2}$ and shape $\frac{k}{2}$.

This distribution has mean k and variance $2k$.

Since the sum of k squared standard normal distributions are distributed as χ^2_k, we can sample from the χ^2 distribution by repeatedly sampling from a normal distribution and summing the squares of observations. Since the inverse of F doesn't have a closed form, this is a computationally convenient method.

Programs

We can sample from the gamma distribution with `chi_square_sampler`:

```r
k <- 1
n <- 200

chi_samples <- chi_square_sampler(n, k)
head(chi_samples)
```

```
## [1] 1.16089567 1.48298239 0.87845621 1.13166616 0.02241891 1.12746920
```

R also has a built-in gamma distribution with pdf `dchisq`, which can be sampled with `rchisq`. We can test that this matches our implementation using a one-sample Kolmogorov-Smirnov test:

```r
ks_results <- ks.test(chi_samples, "dchisq", df = k)
ks_results[["p.value"]]
```

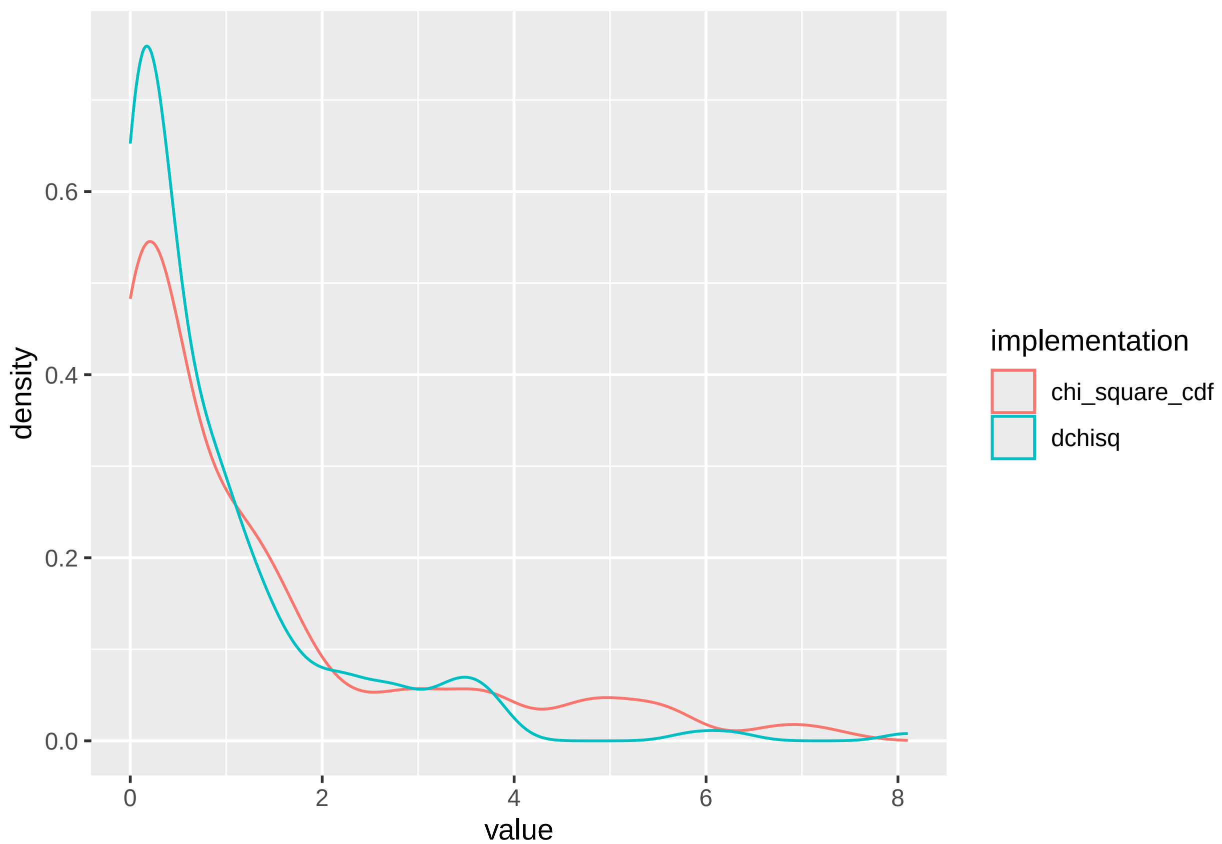
```
## [1] 0
```

In this instance, since $p < 0.05$, we conclude that the distributions match.

We can also plot the distributions against each other:

```r
dat <- data.frame(implementation =
                    factor(rep(c("chi_square_cdf", "dchisq"), each = n)),
                  value = c(chi_samples,
                            rchisq(n, df = k)))

ggplot(dat, aes(x = value, colour = implementation)) + geom_density()
```



Here, we have some noise from our relatively low number of samples, but the distributions clearly have the same shape.

Problem 13

Write a program to generate a random sample of size n from each of the following distributions:

- Chi-square with 1 degree of freedom (χ^2_1);
- Chi-square with 5 degrees of freedom (χ^2_5);
- Chi-square with 40 degrees of freedom (χ^2_{40}).

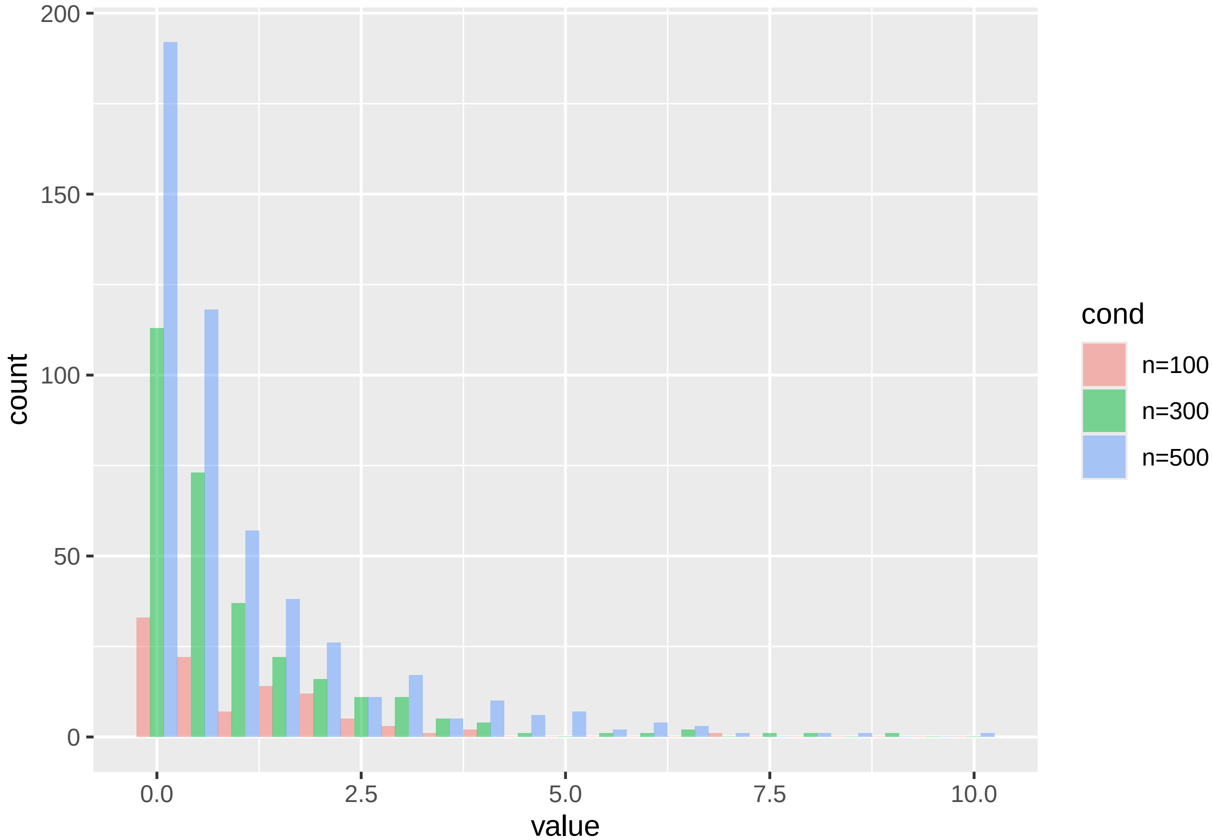
Run your program for $n = 100, 300, 500$ and include a histogram in each case. How do these histograms change in shape as you change the degrees of freedom?

Solution

For the 1 degree of freedom case, we get the following histograms:

```
dat11 <- data.frame(cond = factor(rep("n=100", each = 100)),
                    value = chi_square_sampler(100, 1))
dat12 <- data.frame(cond = factor(rep("n=300", each = 300)),
                    value = chi_square_sampler(300, 1))
dat13 <- data.frame(cond = factor(rep("n=500", each = 500)),
                    value = chi_square_sampler(500, 1))

ggplot(rbind(dat11, dat12, dat13),
       aes(x = value, fill = cond)) +
  geom_histogram(binwidth = 0.5, alpha = 0.5, position = "dodge")
```

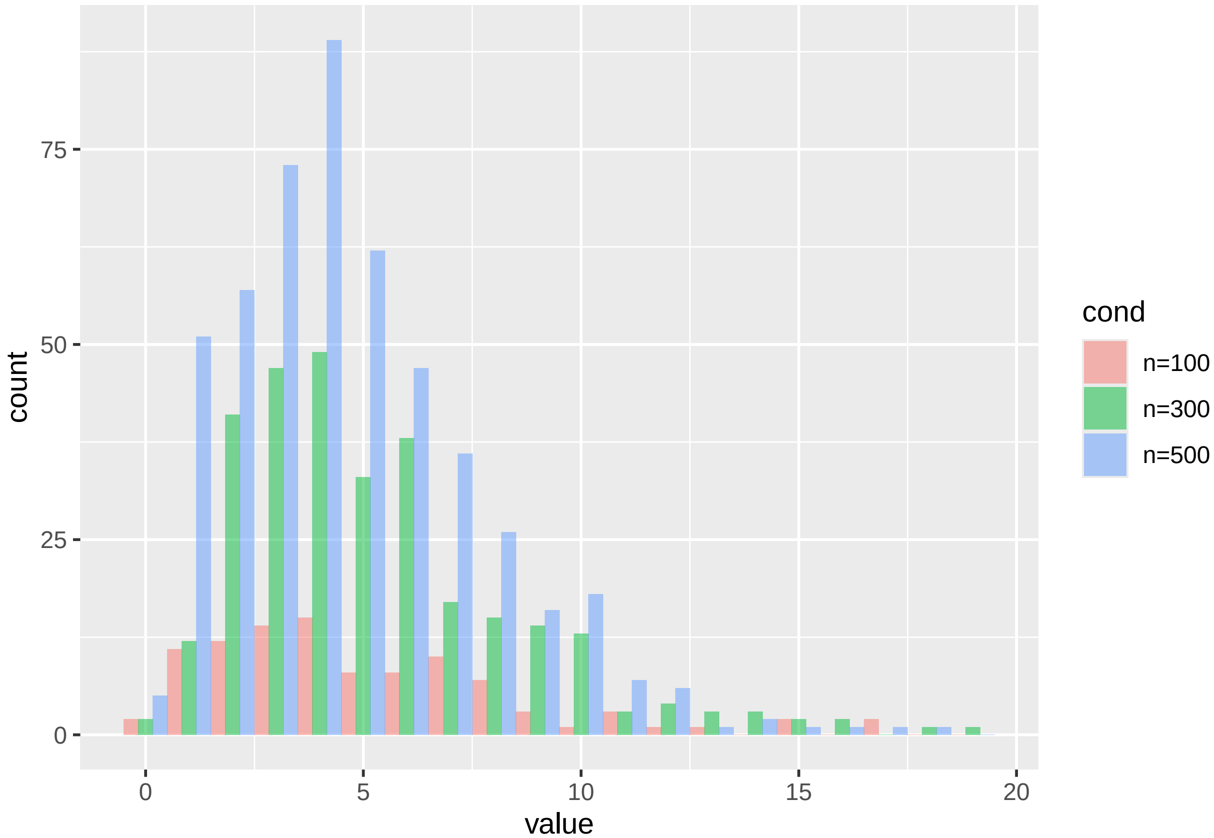


For the 5 degrees of freedom case, we get the following histogram:

```
dat21 <- data.frame(cond = factor(rep("n=100", each = 100)),
                    value = chi_square_sampler(100, 5))
dat22 <- data.frame(cond = factor(rep("n=300", each = 300)),
                    value = chi_square_sampler(300, 5))
dat23 <- data.frame(cond = factor(rep("n=500", each = 500)),
                    value = chi_square_sampler(500, 5))

ggplot(rbind(dat21, dat22, dat23),
```

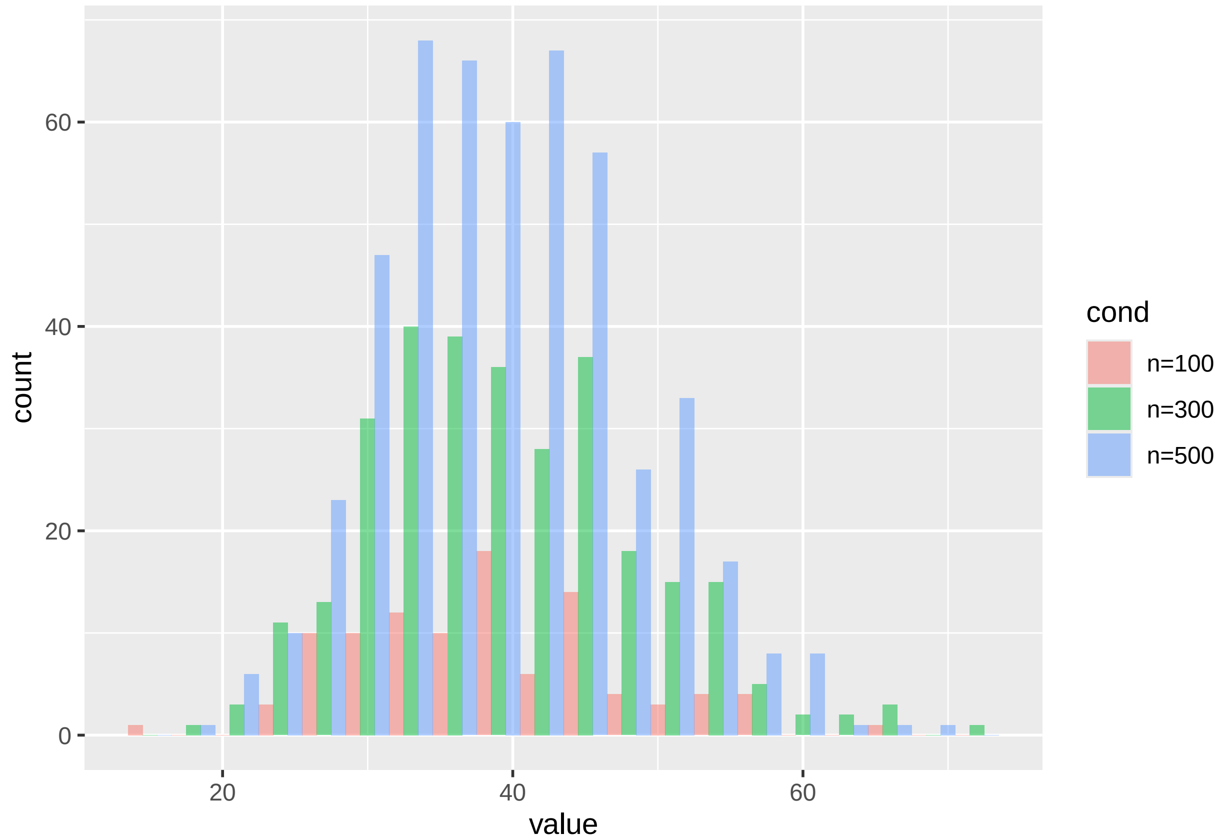
```
       aes(x = value, fill = cond)) +
  geom_histogram(binwidth = 1, alpha = 0.5, position = "dodge")
```



And for the 40 degrees of freedom case, we get the following histogram:

```
dat31 <- data.frame(cond = factor(rep("n=100", each = 100)),
                    value = chi_square_sampler(100, 40))
dat32 <- data.frame(cond = factor(rep("n=300", each = 300)),
                    value = chi_square_sampler(300, 40))
dat33 <- data.frame(cond = factor(rep("n=500", each = 100)),
                    value = chi_square_sampler(500, 40))

ggplot(rbind(dat31, dat32, dat33),
       aes(x = value, fill = cond)) +
  geom_histogram(binwidth = 3, alpha = 0.5, position = "dodge")
```



Note that, as the degrees of freedom increase, the peak of the distribution shifts further right. Indeed, since the skew of the χ^2_k distribution is $\sqrt{\frac{8}{k}}$, we should expect the left tail to grow as k increases. For high values of k, the skew will be approximately 0, so the distribution will have balanced left and right tails.