2 Monte Carlo Methods for Hypothesis Tests

There are two aspects of hypothesis tests that we will investigate through the use of Monte Carlo methods: Type I error and Power.

Example 2.1 Assume we want to test the following hypotheses

$$H_0 : \mu = 5$$
$$H_a : \mu > 5$$

with the test statistic

$$T^* = \frac{\bar{x} - 5}{s/\sqrt{n}}.$$

This leads to the following decision rule:

What are we assuming about X?

2.1 Types of Errors

Type I error:

Type II error:

Usually we set $\alpha = 0.05$ or 0.10, and choose a sample size such that power $= 1 - \beta \geq 0.80$.

For simple cases, we can find formulas for α and β.

2.2 MC Estimator of α

Assume $X_1, \ldots, X_n \sim F(\theta_0)$ (i.e., assume H_0 is true).

Then, we have the following hypothesis test –

$$H_0 : \theta = \theta_0$$
$$H_a : \theta > \theta_0$$

and the statistics T^*, which is a test statistic computed from data. Then we **reject** H_0 if $T^* >$ the critical value from the distribution of the test statistic.

This leads to the following algorithm to estimate the Type I error of the test (α)

Your Turn

Example 2.2 (Pearson's moment coefficient of skewness) Let $X \sim F$ where $E(X) = \mu$ and $Var(X) = \sigma^2$. Let

$$\sqrt{\beta_1} = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right].$$

Then for a

- symmetric distribution, $\sqrt{\beta_1} = 0$,
- positively skewed distribution, $\sqrt{\beta_1} > 0$, and
- negatively skewed distribution, $\sqrt{\beta_1} < 0$.

The following is an estimator for skewness

$$\sqrt{b_1} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{3/2}}.$$

It can be shown by Statistical theory that if $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, then as $n \to \infty$,

$$\sqrt{b_1} \stackrel{.}{\sim} N\left(0, \frac{6}{n}\right).$$

Thus we can test the following hypothesis

$$H_0 : \sqrt{\beta_1} = 0$$
$$H_a : \sqrt{\beta_1} \neq 0$$

by comparing $\dfrac{\sqrt{b_1}}{\sqrt{\frac{6}{n}}}$ to a critical value from a $N(0,1)$ distribution.

In practice, convergence of $\sqrt{b_1}$ to a $N\left(0, \frac{6}{n}\right)$ is slow.
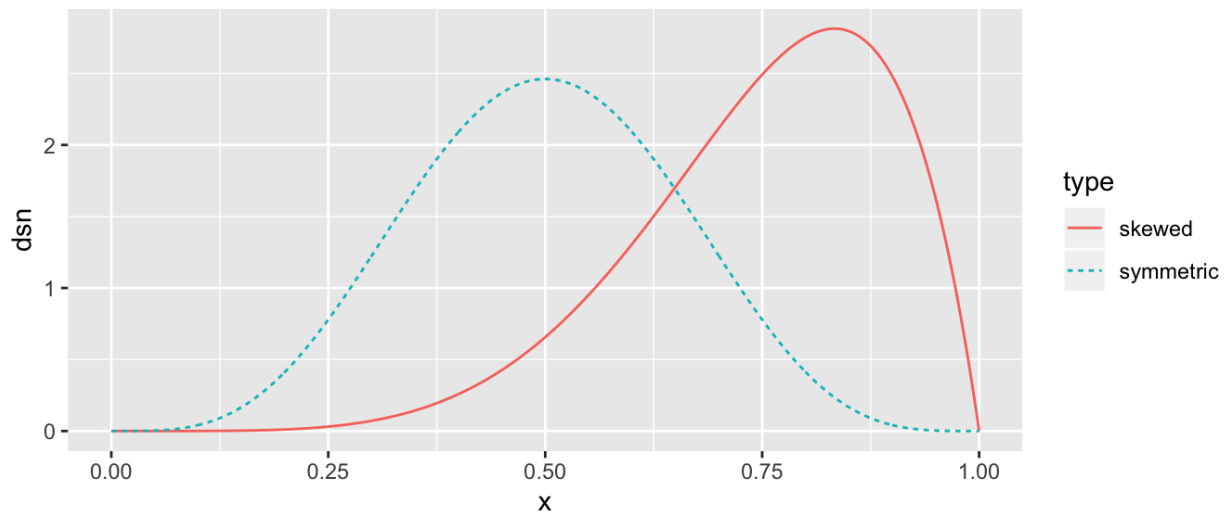
We want to assess $P(\text{Type I error})$ for $\alpha = 0.05$ for $n = 10, 20, 30, 50, 100, 500$.

```r
library(tidyverse)

# compare a symmetric and skewed distribution
data.frame(x = seq(0, 1, length.out = 1000)) %>%
  mutate(skewed = dbeta(x, 6, 2),
         symmetric = dbeta(x, 5, 5)) %>%
  gather(type, dsn, -x) %>%
  ggplot() +
  geom_line(aes(x, dsn, colour = type, lty = type))
```



```r
## write a skewness function based on a sample x
skew <- function(x) {

}

## check skewness of some samples
n <- 100
a1 <- rbeta(n, 6, 2)
a2 <- rbeta(n, 2, 6)

## two symmetric samples
b1 <- rnorm(100)
b2 <- rnorm(100)

## fill in the skewness values
ggplot() + geom_histogram(aes(a1)) + xlab("Beta(6, 2)") +
  ggtitle(paste("Skewness = "))
```
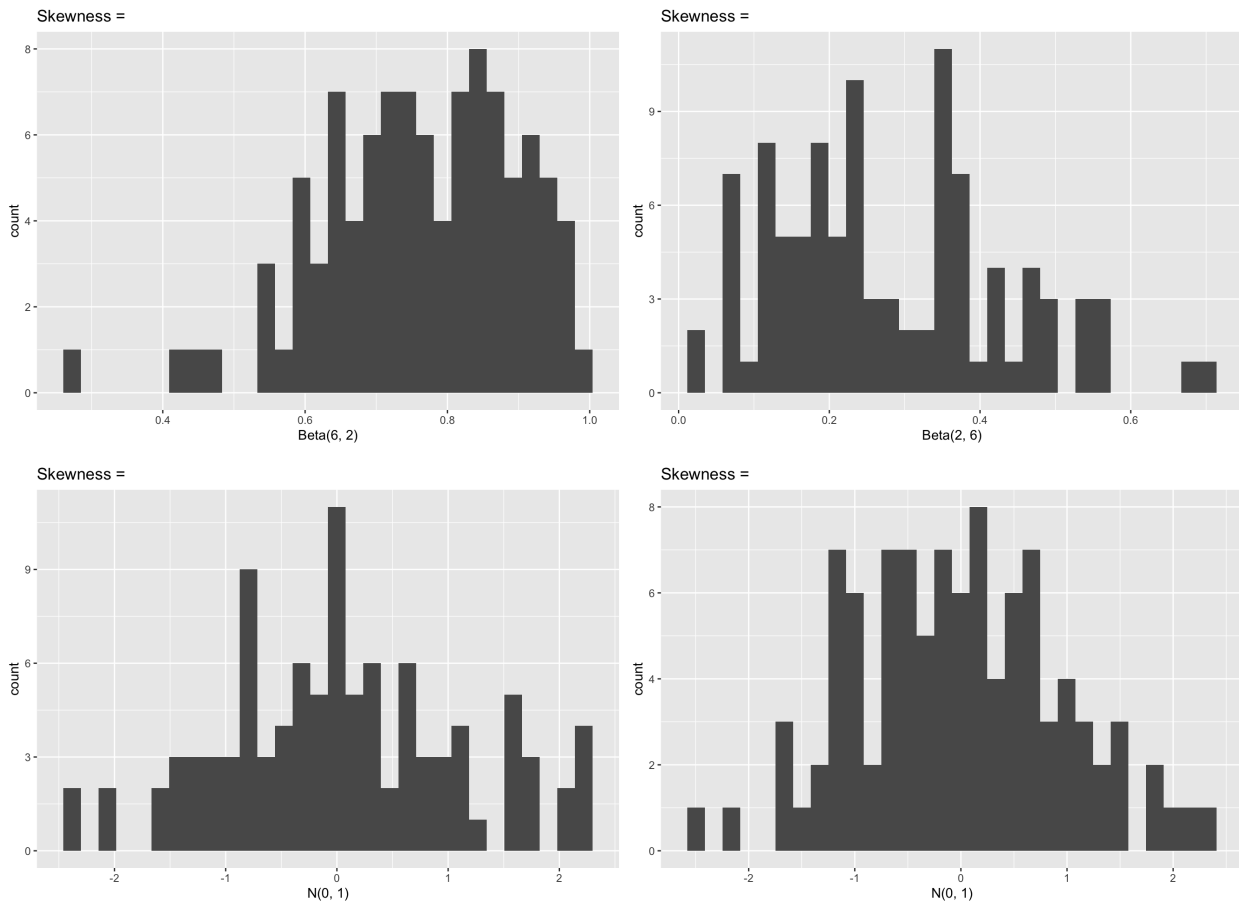
```r
ggplot() + geom_histogram(aes(a2)) + xlab("Beta(2, 6)") +
  ggtitle(paste("Skewness = "))
ggplot() + geom_histogram(aes(b1)) + xlab("N(0, 1)") +
  ggtitle(paste("Skewness = "))
ggplot() + geom_histogram(aes(b2)) + xlab("N(0, 1)") +
  ggtitle(paste("Skewness = "))
```



```r
## Assess the P(Type I Error) for alpha = .05, n = 10, 20, 30, 50,
   100, 500
```

Example 2.3 (Pearson's moment coefficient of skewness with variance correction) One way to improve performance of this statistic is to adjust the variance for small samples. It can be shown that

$$Var(\sqrt{b_1}) = \frac{6(n-2)}{(n+1)(n+3)}.$$

Assess the Type I error rate of a skewness test using the finite sample correction variance.