QBS 120 - Problem Set 6

Rob Frost

Grading: Problems 3 and 5; 4 pts for problem 3, 6 pts for problem 5; see details below for each problem

1. **(Based on Rice 9.1) A coin is thrown independently 10 times to test the hypothesis that the probability of a heads is 1/2 vs the alternative that the probability is not 1/2. The test rejects if the number of observed heads is 0 or 10.**

   We'll model the coin throws as independent Bernoulli RVs so the number of heads is a binomial RV with n=10. The value of p is given by the following simple null and composite alternative hypotheses:

   $$H_0 : p = 0.5$$
   $$H_A : p \neq 0.5$$

   Let X be the number of heads.

   (a) **What is the significance level of the test?**

   Remember that the significance level, α, is the probability of rejecting H_0 when H_0 is true:

   $$\alpha = P(X = 0 \cup X = 10 | H_0)$$
   $$= P(X = 0 \cup X = 10 | p = 0.5)$$
   $$= P(X = 0 | p = 0.5) + P(X = 10 | p = 0.5)$$

   These probabilities can be computed using the binomial PMF:

   ```
   > (alpha=dbinom(0, 10, 0.5) + dbinom(10, 10, 0.5))
   [1] 0.001953125
   ```

   (b) **If the probability of heads is 0.1, what is the power of the test?**

   We are given the specific alternative:

   $$H_A : p = 0.1$$

   This allows us to the compute the power (or the probability of rejecting H_0 if H_A is true):

   $$power = P(X = 0 \cup X = 10 | H_A)$$
   $$= P(X = 0 \cup X = 10 | p = 0.1)$$
   $$= P(X = 0 | p = 0.1) + P(X = 10 | p = 0.1)$$

   These probabilities can be computed using the binomial PMF:

```
> (power=dbinom(0, 10, 0.1) + dbinom(10, 10, 0.1))
[1] 0.3486784
```

(c) **If the test instead rejects if the number of observed heads is ≤ 1 or ≥ 9, what is the significance level?**

To answer this, we just modify the approach in a) to include $X = 1$ and $X = 9$:

$$\begin{aligned}
\alpha &= P(X = 0 \cup X = 1 \cup X = 9 \cup X = 10 | H_0) \\
&= P(X = 0 \cup X = 1 \cup X = 9 \cup X = 10 | p = 0.5) \\
&= P(X = 0 | p = 0.5) + P(X = 1 | p = 0.5) + P(X = 9 | p = 0.5) + P(X = 10 | p = 0.5)
\end{aligned}$$

These probabilities can be computed using the binomial PMF:

```
> (alpha=dbinom(0, 10, 0.5)
+          + dbinom(1, 10, 0.5)
+          + dbinom(9, 10, 0.5)
+          + dbinom(10, 10, 0.5))
[1] 0.02148438
```

As expected, the significance level is larger given the increased rejection region.

2. **(Based on Rice 9.2) Which of the following hypotheses are simple, and which are composite? Justify your answers.**

(a) **X follows a uniform distribution on [0,1].**

Simple. The parameters a and b of a uniform distribution have exact values (0 and 1).

(b) **A die is unbiased.**

Simple. We know that the probability of each of the six outcomes is $1/6$ (i.e., a discrete uniform distribution).

(c) **X follows a normal distribution with mean 0 and variance $\sigma^2 > 10$.**

Composite, σ^2 can take an infinite number of values.

(d) **X follows a normal distribution with mean $\mu = 0$.**

Composite, σ^2 can take any value > 0.

3. **(Based on Rice 9.5) True or false and state why:**

Grading: 4 pts total. Each part is worth 0.5 pts. To get credit need to have the correct answer and some plausible justification (exact wording is not required).

(a) **The significance level of a statistical test is equal to the probability that the null hypothesis is true.**

False. The significance level is the type I error rate or the probability of rejection H_0 when H_0 is true.

(b) **If the significance level of a test is decreased, the power would be expected to increase.**

False. The exact opposite happens: decreasing the significance level lowers the power.

(c) **If a test is rejected at the significance level α, the probability that the null hypothesis is true equals α.**

False. The significance level is the probability of a rejecting the test assuming H_0 is true. It is a conditional probability. It tells us nothing about the marginal probability of H_0. This type of probability is used as a prior in Bayesian analysis.

(d) **The probability that the null hypothesis is falsely rejected is equal to the power of the test.**

False. The probability of rejecting the null when the null is true is the type I error rate, which does not have a direct relationship to the power. The power is the chance of rejecting the null when the alternate is true.

(e) **A type I error occurs when the test statistic falls in the rejection region of the test.**

False. A type I error occurs when the null is rejected (i.e., the test statistic falls in the rejection region) **and** the null hypothesis is true. Both conditions are needed for a type I error. If the null is false and the test statistic is in the rejection region, no error was made.

(f) **A type II error is more serious than a type I error.**

False. The type of error that is most serious is very context dependent. For exploratory/hypothesis generating experiments, type II errors are more serious (i.e., we are OK with some false positives as long as nothing interesting is missed). For DNA matching in criminal cases, type I errors are much more serious (i.e., do not want to falsely implicate someone).

(g) **The power of a test is determined by the null distribution of the test statistic.**

False. Power is computed under the alternate distribution (power $= \mathrm{P}(\text{rejecting } H_0 | H_A)$).

(h) **The likelihood ratio is a random variable.**

True! The likelihood ratio is a function of the data and so will change each time the experiment is performed.

4. **(Based on Rice 9.30) Suppose that the null hypothesis is true, that the distribution of the test statistic, T say, is continuous with cdf F and that the test rejects for large values of T. Let V denote the p-value of the test.**

Grading: 6 pts total, see details for each part

(a) **Show that $V = 1 - F(T)$.**

Grading: 1 pt; have to provide a plausible answer, exact wording not important

By definition, the p-value is the smallest significance level α for which the null hypothesis is rejected. It is equivalently defined as the probability of encountering a test statistic equal to or greater in magnitude than the observed value under H_0. We'll use this second definition. Since the test rejects for large values of T, the p-value is the probability of encountering a value larger than T under H_0. Given that the CDF under the null is $F()$, this probability is:

$$P(X \geq T | H_0) = 1 - F(T)$$

The p-value V therefore equals $1 - F(T)$.

(b) **Conclude that the null distribution of V is uniform.**

Grading: 2 pts; 1 pt to show that $P(V \leq c) = c$, 1 pt for a plausible explanation

Sine $F(T)$ is a valid CDF, we know it has a standard uniform distribution when T has the distribution F(). So, under $H_0 : F(T) \sim U(0,1)$. V is therefore a a linear function of a uniform under the null, which is also uniform; in this case a U(0,1) since $1 - U(0,1) \sim U(0,1)$. In general, p-values always have a $U(0,1)$ distribution under the null.

We can show that $V \sim U(0,1)$ as follows:

$$
\begin{aligned}
P(V \leq c) &= P(1 - F(T) \leq c) && \text{def of CDF of V} \\
&= P(F(T) \geq 1 - c) \\
&= P(T \geq F^{-1}(1 - c)) \\
&= c && \text{by definition of inverse CDF}
\end{aligned}
$$

V therefore has a U(0,1) distribution

(c) **If the null hypothesis is true, what is the probability that the p-value is greater than .1?**

Grading: 1 pt

Since the p-value under the null is distributed uniformly between 0 and 1, this probability will be .9 by definition.

(d) **Show that the test that rejects if $V < \alpha$ has significance level α.**

Grading: 2 pts; 1 pt to correctly define the significance level, 1pt to show the desired result

Significance level $= P(V < \alpha | H_0)$:

$$
\begin{aligned}
P(V < \alpha | H_0) &= \int_0^\alpha dv, \text{ pdf of U(0,1)} \\
P(V < \alpha | H_0) &= v |_0^\alpha \\
P(V < \alpha | H_0) &= \alpha
\end{aligned}
$$

The significance level of the test that rejects if $V < \alpha$ is therefore α

5. **(Based on Rice 9.42) Nylon bars were tested for brittleness. Each of the 280 bars was molded under similar conditions and was tested in five places. Assuming that each bar has uniform composition, the number of breaks on a given bar should be binomially distributed with five trials and an unknown probability of p of failure. If the bars are all of the same uniform strength, p should be the same for all of them; if they are of different strengths, p should vary from bar to bar. Thus, the H_0 is that the p's are all equal. The following table summarizes the outcome of the experiment:**

Breaks/Bar	Frequency
0	157
1	69
2	35
3	17
4	1
5	1

```
>          (sum = 69+35*2+17*3+4+5)
```

[1] 199

```
>          (n = 157+69+35+17+1+1)
```

[1] 280

(a) **Under the given assumption, the data in the table consist of 280 observations of independent binomial random variables. Find the mle of p.**

For 280 binomial RVs with $n = 5$ and $p = p$, the MLE of p can be computed as follows:

$$L(p) = \prod_{i=1}^{280} \binom{5}{X_i} p^{X_i}(1-p)^{5-X_i}$$

$$log(L(p)) = l(p) = log(\prod_{i=1}^{280} \binom{5}{X_i} p^{X_i}(1-p)^{5-X_i})$$

$$l(p) = \sum_{i=1}^{280}(log(\binom{5}{X_i})) + X_i log(p) + (5 - X_i)log(1-p))$$

$$\frac{\delta}{\delta p}l(p) = 1/p \sum_{i=1}^{280} X_i - 1/(1-p)\sum_{i=1}^{280}(5 - X_i)$$

$$0 = 199/p - 1201/(1-p)$$

$$0 = (1-p)199 - 1201p$$

$$1400p = 199$$

$$p = 0.142$$

(b) **Pooling the last three cells, test the agreement of the observed frequency distribution with the binomial distribution using Pearson's chi-square test.**

The Pearson X^2 statistic is:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

We first need to compute the expected counts without pooling. We'll then pool the last 3 cells and calculate X^2.

Expected counts given MLE:

```
>           observed = c(157,69,35,17,1,1)
>           breaks = 0:5
>           probs = sapply(breaks, function(x) {
+                   dbinom(x, size=5, prob=.142)
+           })
>           expected = sapply(probs, function(x) {
+                   n*x
+           })
>           (counts = data.frame(breaks, probs, observed, expected))
  breaks         probs observed     expected
1      0 4.649823e-01      157 130.19504936
2      1 3.847756e-01       69 107.73716206
3      2 1.273616e-01       35  35.66125178
4      3 2.107850e-02       17   5.90197873
5      4 1.744258e-03        1   0.48839218
6      5 5.773534e-05        1   0.01616589
```

Now merge the last 3 cells:

```
>           (merged  = data.frame(breaks=0:3,
+                   probs = c(probs[1:3], sum(probs[4:6])),
+                   observed=c(observed[1:3], sum(observed[4:6])),
+                   expected=c(expected[1:3], sum(expected[4:6])))))
  breaks       probs observed   expected
1      0 0.46498232      157 130.195049
2      1 0.38477558       69 107.737162
3      2 0.12736161       35  35.661252
4      3 0.02288049       19   6.406537
```

Pearson's chi-square statistic with 4-1-1=2 dof (number of cells - number of estimated parameters - 1 since cell probs must sum to 1).

```
>           pearson = sum(sapply(1:4, function(x) {
+                   (merged$observed[x] - merged$expected[x])^2/merged$expected[x]}))
>           pearson
[1] 44.21422
>           pchisq(pearson, df=2, lower.tail=F)
[1] 2.506131e-10
```

We can therefore confidently reject the null hypothesis that the data conform to a binomal distribution.

6. **In this problem you will develop and apply a test for normality based on the coefficient of skewness as defined in Section 9.9 of Rice.**

   (a) **Write an R function to compute the coefficient of skewness, b_1.**

   Although you can certainly find existing R implementations, we'll implement from scratch. Note: to match the implementation of skewness in the R moments package, we'll compute the standard deviation using the biased variance instead of the unbiased variance as is done in the standard R sd() function.

6

```
> biased.sd = function(x) {
+        biased.var = mean((x-mean(x))^2)
+        return (sqrt(biased.var))
+ }
> coef.of.skewness = function(x) {
+        b.1 = mean((x - mean(x))^3)/biased.sd(x)^3
+        return (b.1)
+ }
```

To make sure we implemented correctly, let's compare to an existing implementation:

```
> library(moments)
> x = rnorm(10)
> coef.of.skewness(x)

[1] -0.3067451

> skewness(x)

[1] -0.3067451
```

(b) **Use simulation to approximate the sampling distribution of b_1 when the data is modeled by 100 independent $\mathcal{N}(0, 1)$ RVs.**

First, let's simulate 10,000 data sets and compute b_1 on each:

```
> B=10000
> n=100
> sim.data = matrix(rnorm(B*n), nrow=B, ncol=n)
> sim.b.1 = apply(sim.data, 1, coef.of.skewness)
```

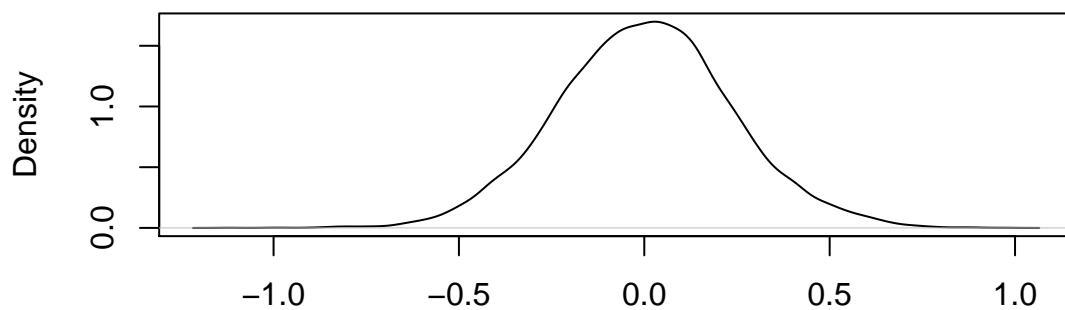Let's visualize the sampling distribution of b_1 when the data is iid standard normal:

```
> par(mfrow=c(2,1))
> hist(sim.b.1)
> plot(density(sim.b.1))
```
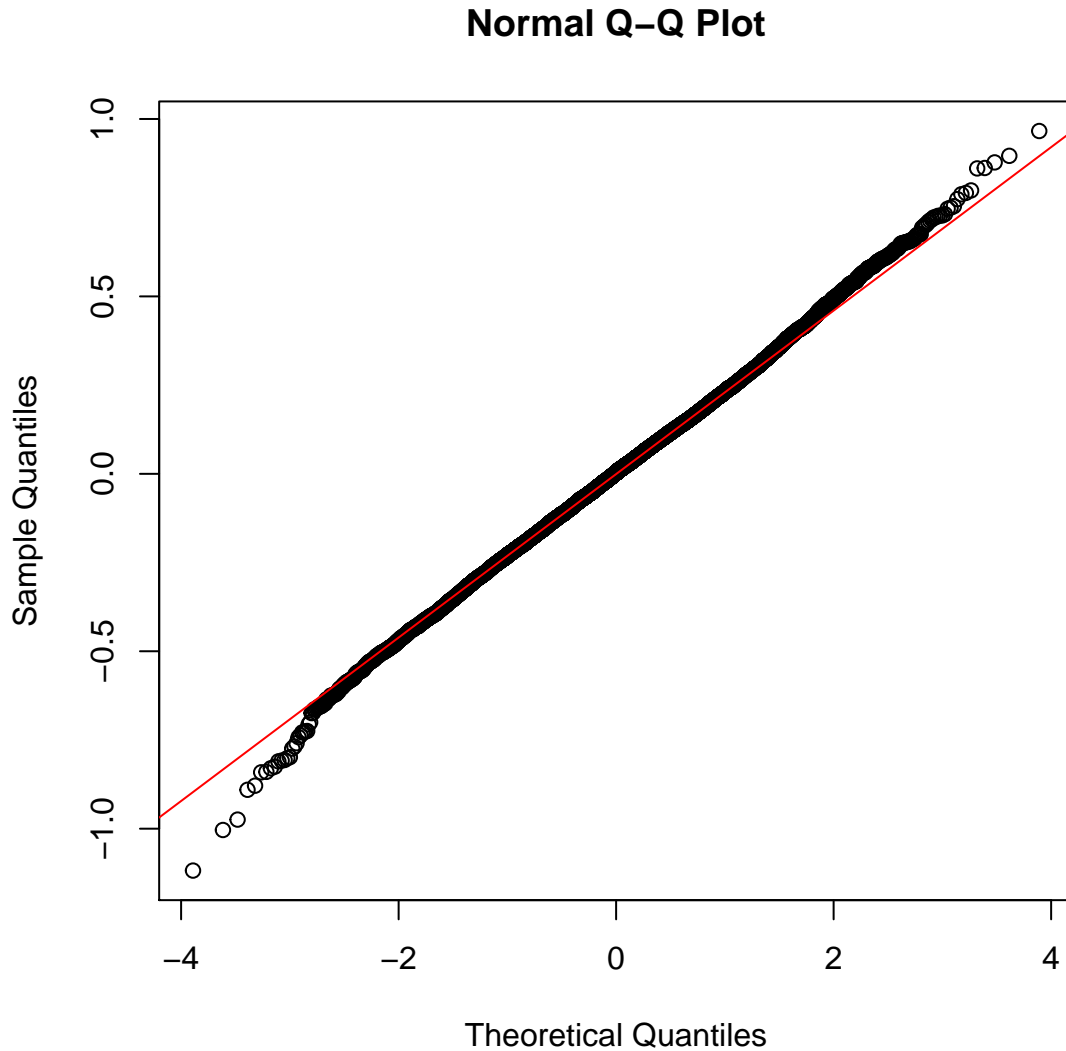
Histogram of sim.b.1
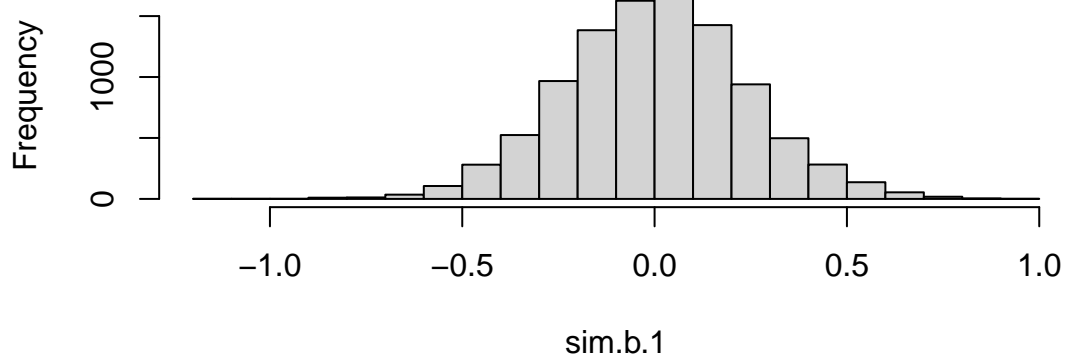
density.default(x = sim.b.1)

N = 10000   Bandwidth = 0.03307

The sampling distribution in this case appears approximately normal with mean 0. Let's test this by generating a probability plot relative to the standard normal quantiles:

```
> qqnorm(sim.b.1)
> qqline(sim.b.1, distribution=qnorm, col="red")
```

Normal Q−Q Plot

A standard normal appears to be a good match.

(c) **Test 1000 data sets each containing 100 independent $\mathcal{N}(0,1)$ RVs for normality. For this test H_0 is that the data are iid $\mathcal{N}(0,1)$ and H_A is that the data are not iid $\mathcal{N}(0,1)$. Use the sampling distribution of b_1 to compute an approximate p-value for each test.**

We'll compute the p-values using the quantiles of the simulated b_1 values. Since the p-value is the probability of finding a b_1 value equal to or greater in magnitude (i.e., larger absolute value) under H_0, we will compute the p-value as follows:

- Find the largest quantile of the simulated null distribution that is $\leq b_1$. Call the quantile of this α_{low}.
- Find the smallest quantile of the simulated null distribution that is $\geq b_1$. Call 1 minus the quantile of this α_{hi}.
- The p-value for a two-sided test is $2 * min(\alpha_{low}, \alpha_{hi})$.

```
> test.data = matrix(rnorm(1000*100), nrow=1000, ncol=100)
```

```
> test.b.1 = apply(test.data, 1, coef.of.skewness)
> ranked.sim.b.1 = sort(sim.b.1)
> simPVal = function(x, ranked.sim.values) {
+       n = length(ranked.sim.values)
+       smaller.vals =which(ranked.sim.values <= x)
+       if (length(smaller.vals) == 0) {
+               alpha.low = 0
+       } else{
+               alpha.low = length(smaller.vals)/n
+       }
+       larger.vals =which(ranked.sim.values >= x)
+       if (length(larger.vals) == 0) {
+               alpha.hi = 0
+       } else{
+               alpha.hi = length(larger.vals)/n
+       }
+       p.val = 2*min(alpha.low, alpha.hi)
+       return (p.val)
+ }
> p.values = sapply(test.b.1, function(x) simPVal(x, ranked.sim.b.1))
> p.values[1:10]
 [1] 0.9906 0.6330 0.8162 0.7054 0.0720 0.7212 0.9610 0.5924 0.6952 0.0108
```
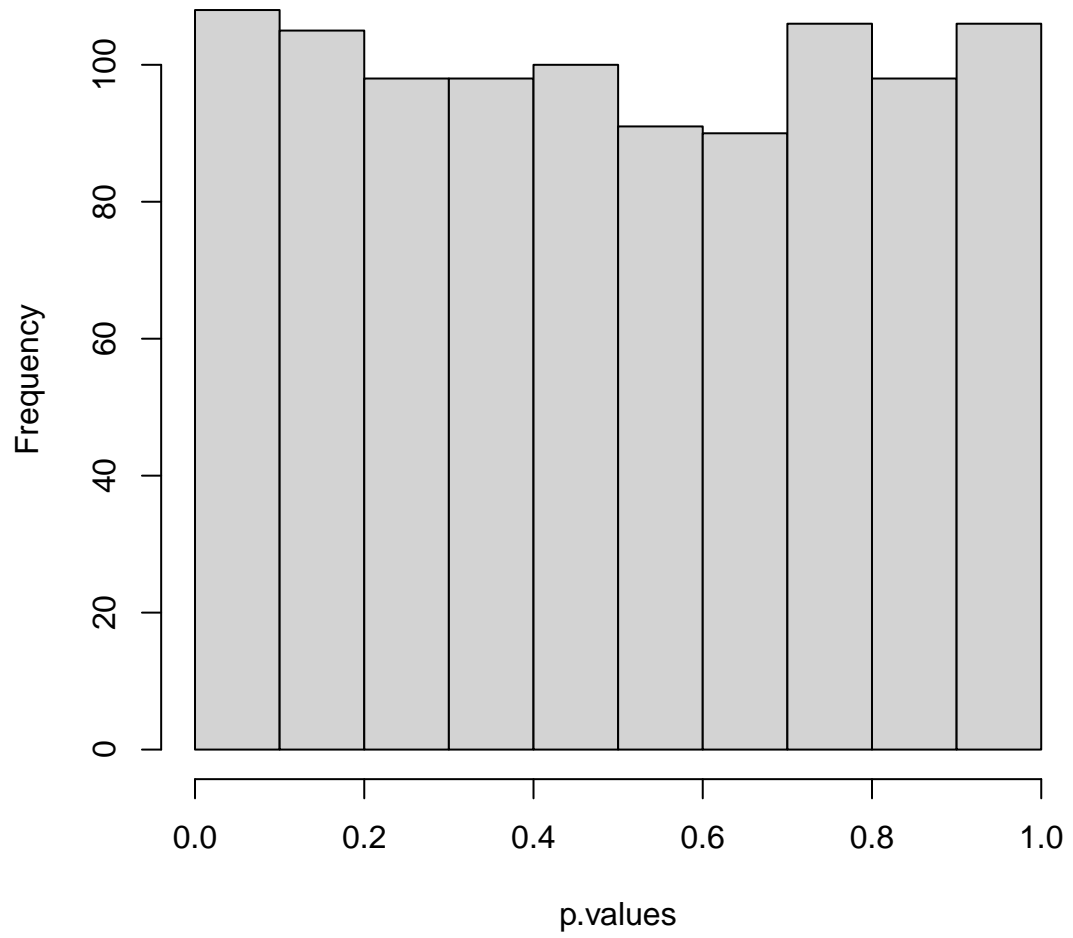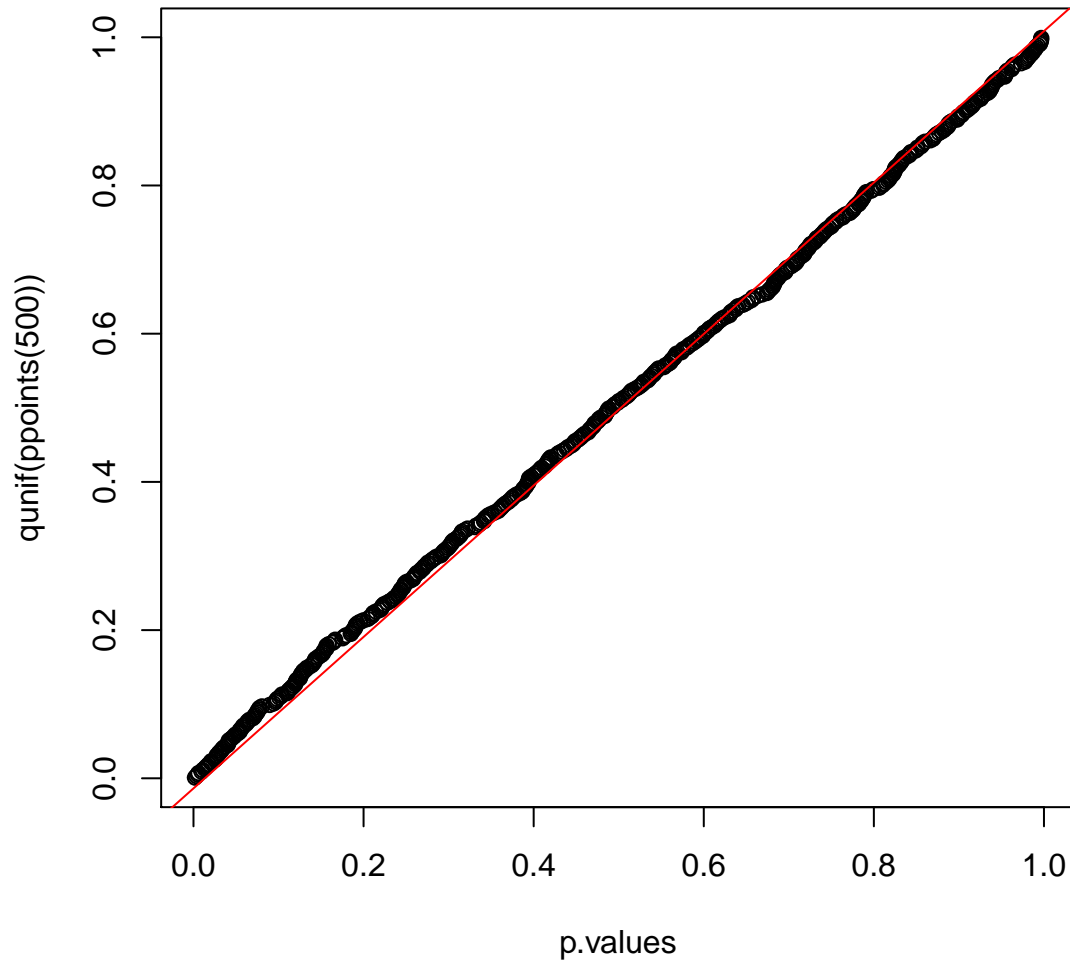
Generate a Q-Q plot of these p-values relative to a theoretical $U(0, 1)$ distribution.

Let's visualize these p-values using both a histogram and a probability plot vs. a standard uniform:

Histogram of p.values

Does the shape of this plot match your expectations? Explain.

We expect that if H_0 is true, the p-values will have a U(0,1) distribution. The results match our expectation.

(d) **Test 1000 data sets each containing 100 independent Poisson RVs with $\lambda = 1$ for normality. Compute an approximate p-value for each test.**

```
> test.data = matrix(rpois(1000*100, lambda=2), nrow=1000, ncol=100)
> test.b.1 = apply(test.data, 1, coef.of.skewness)
> p.values = sapply(test.b.1, function(x) simPVal(x, ranked.sim.b.1))
> p.values[1:10]
 [1] 0.0000 0.0746 0.0634 0.1528 0.2122 0.0022 0.0660 0.0050 0.0580 0.0102
```

Generate a Q-Q plot of these p-values relative to a theoretical $U(0,1)$ distribution.

We will again visualize these p-values using both a histogram and a probability plot vs. a standard uniform:

Histogram of p.values
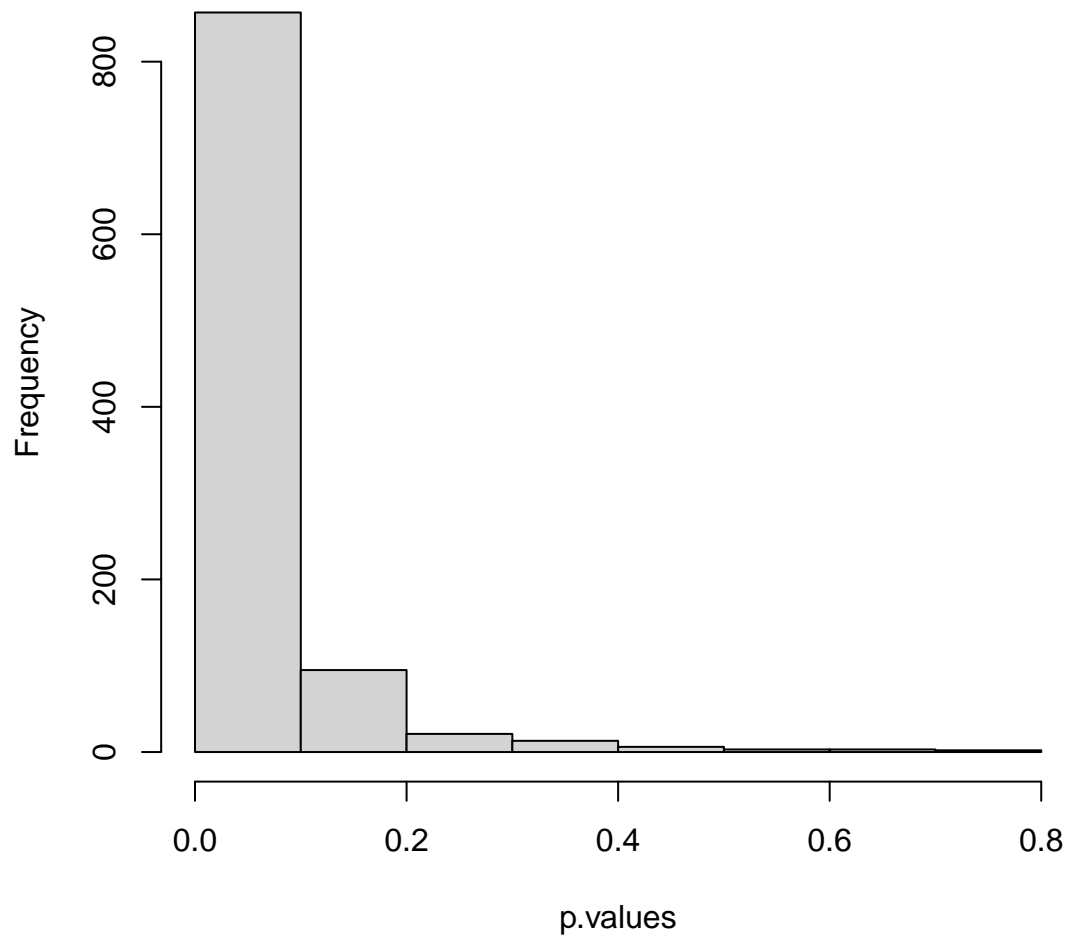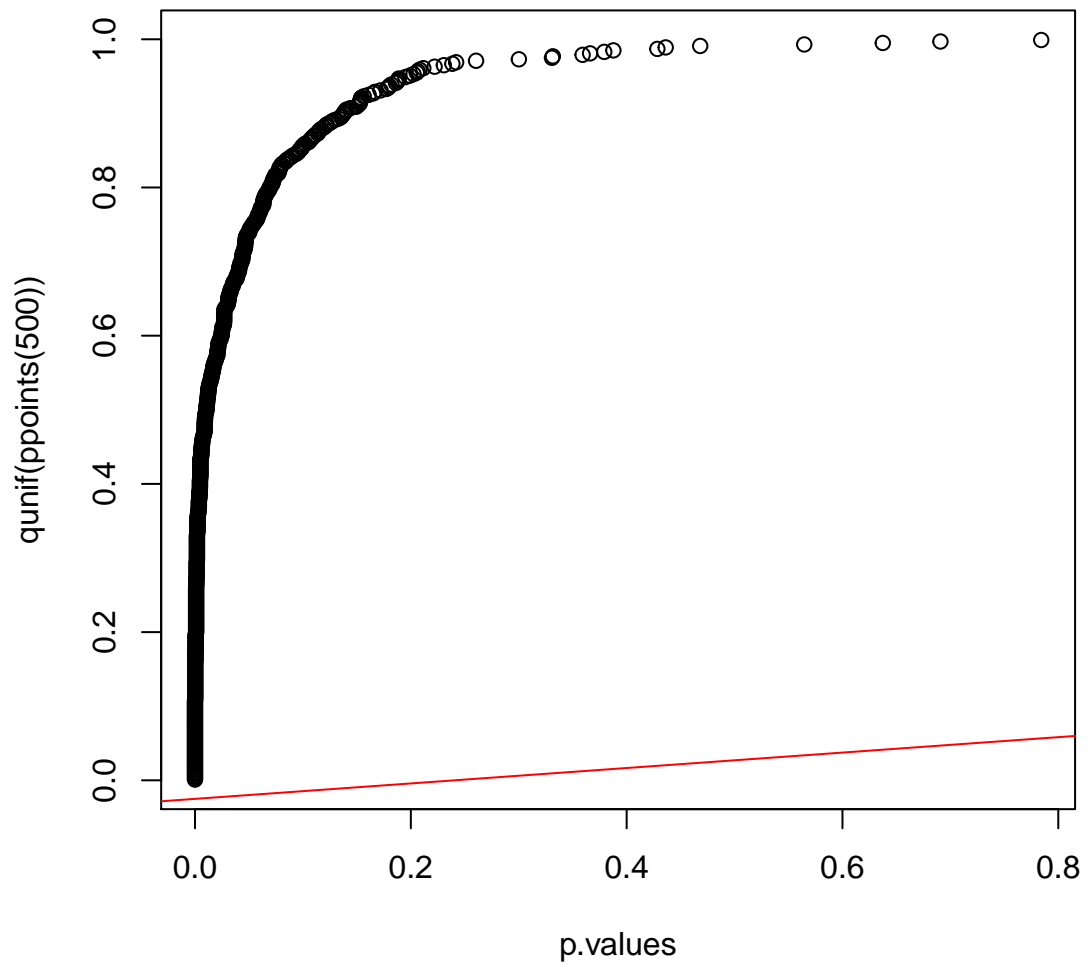
Does the shape of this plot match your expectations? Explain.

In this case, the data are simulated under H_A and the low value of λ means that the normal approximate for the Poisson does not hold well so we would expect the p-values to not have a U(0,1) distribution. In particular, we would expect more small p-values that predicted by U(0,1). The results match our expectations.