All solutions are give below. All code used to generate the requisite simulations, figures, etc. is attached.

1. (a) Recall that Maximum Likelihood Estimators (MLE) are invariant. That is if $\hat{\theta}_{MLE}$ is a maximum likelihood estimate for θ, then for any real valued function $\widehat{g(\theta)}_{MLE} = g(\hat{\theta}_{MLE})$. Let $X_1, X_2, \dots, X_N$ be iid $Poisson(\theta)$. Then the likelihood is given by

$$\mathcal{L}(X_1, \dots, X_N|\theta) = f(X_1, X_2, \dots, X_N|\theta) = \prod_{i=1}^{n} f(X_i|\theta) = \prod_{i=1}^{n} \frac{\theta^{X_i}e^{-\theta}}{X_i!} = \frac{\theta^{\sum_{i=1}^{n} X_i}e^{-n\theta}}{X_1!X_2!\dots X_n!}$$

From here we can find the log-likelihood as

$$\ell(X_1, X_2, \dots, X_n|\theta) = \log(\mathcal{L}(X_1, X_2, \dots, X_n|\theta)) = \sum_{i=1}^{n} X_i \log(\theta) - n\theta - \log(X_1!X_2!\dots X_n!)$$

Taking the derivative and setting to zero, we can find the maximum likelihood estimate of θ.

$$\frac{\partial}{\partial\theta}\left(\sum_{i=1}^{n} X_i \log(\theta) - n\theta - \log(X_1!X_2!\dots X_n!)\right) = \frac{\sum_{i=1}^{n} X_i}{\theta} - n \overset{set}{=} 0$$

Solving for θ, we have $\hat{\theta}_{MLE} = \overline{X}$. Hence, our MLE estimate for $\pi(\theta)$ is given by $\widehat{\pi(\theta)}_{MLE} = e^{-\overline{X}}$.

(b) To estimate the Bias of $e^{-overlineX}$, we must first calculate the expected value of the estimate. We will use a Taylor series expansion around the point $x = \theta$ to aid the approximations.

$$\mathbb{E}(e^{-\overline{X}}) = \mathbb{E}\Big[\sum_{n=0}^{\infty} \frac{(-1)^n e^{-\theta}}{n!}(\overline{X} - \theta)^n\Big] \approx \mathbb{E}\Big[e^{-\theta} - e^{-\theta}(\overline{X} - \theta) + \frac{e^{-\theta}}{2}(\overline{X} - \theta)^2\Big]$$

$$= e^{-\theta} - e^{-\theta}\mathbb{E}(\overline{X} - \theta) + \frac{e^{-\theta}}{2}\mathbb{E}(\overline{X} - \theta)^2 = e^{-\theta} + \frac{e^{-\theta}}{2}\mathbb{E}\Big[\overline{X}^2 - 2\overline{X}\theta + \theta^2\Big]$$

$$= e^{-\theta} + \frac{e^{-\theta}}{2}\Big[\mathbb{V}(\overline{X}) + \mathbb{E}(\overline{X})^2 - 2\theta^2 + \theta^2\Big] = e^{-\theta} + \frac{e^{-\theta}}{2}\Big[\frac{\theta}{n} + \theta^2 - 2\theta^2 + \theta^2\Big] = e^{-\theta} + \frac{\theta e^{-\theta}}{2n}$$

This used the fact that $\overline{X} \sim N(\theta, \frac{\theta}{n})$. Using this we compute the bias of the estimator.

$$Bias(e^{-\overline{X}}) = \mathbb{E}(e^{-\overline{X}}) - e^{-\theta} \approx e^{-\theta} + \frac{\theta e^{-\theta}}{2n} - e^{-\theta} = \frac{\theta e^{-\theta}}{2n}$$

We approximate the variance in a similar fashion. We will use the Taylor series expansion of $e^{-2\overline{X}}$ around θ. We will calculate this expectation first, and the variance to follow.

$$\mathbb{E}(e^{-2\overline{X}}) = \mathbb{E}\Big[\sum_{n=0}^{\infty} \frac{(-2)^n e^{-\theta}}{n!}(\overline{X}-\theta)^n\Big] \approx \mathbb{E}\Big[e^{-\theta} - 2e^{-\theta}(\overline{X}-\theta) + 4\frac{e^{-\theta}}{2}(\overline{X}-\theta)^2\Big]$$

$$= e^{-\theta} - 2e^{-\theta}\mathbb{E}(\overline{X}-\theta) + 2e^{-\theta}\mathbb{E}(\overline{X}-\theta)^2 = e^{-\theta} + \frac{e^{-\theta}}{2}\mathbb{E}\Big[\overline{X}^2 - 2\overline{X}\theta + \theta^2\Big]$$

$$= e^{-\theta} + 2e^{-\theta}\Big[\mathbb{V}(\overline{X}) + \mathbb{E}(\overline{X})^2 - 2\theta^2 + \theta^2\Big] = e^{-\theta} + 2e^{-\theta}\Big[\frac{\theta}{n} + \theta^2 - 2\theta^2 + \theta^2\Big] = e^{-\theta} + \frac{2\theta e^{-\theta}}{n}$$

From here, we see that the variance is approximately

$$\mathbb{V}(e^{-\overline{X}}) = \mathbb{E}(e^{-2\overline{X}}) - \Big[\mathbb{E}(e^{-\overline{X}})\Big]^2 \approx e^{-\theta} + \frac{2\theta e^{-\theta}}{n} - \Big[e^{-\theta} + \frac{\theta e^{-\theta}}{2n}\Big]^2$$

(c) For $n = 100$ and $\theta = 5$ gives $\widehat{Bias}(e^{-\overline{X}}) = 0.0001685$ and $\widehat{Var}(e^{-\overline{X}}) = 0.007364$.

(d) We collected 1000 observations of our estimator where each observation was computed after sampling 100 data points from a Poisson random variable with mean 5. With these 1000 observations of our estimate $\hat{\pi}$, we then calculated the empirical bias for each one of these estimators. The mean of these empirical biases was found to be 0.000218. We then calculated the variance of these 1000 observations and found the sample variance to be 2×10^{-6}.

Our approximation comes close to the empirical estimate of the bias, while our approximated variance is order of magnitudes greater than the empirical variance.

(e) Using the same process as above, for $n = 50$ and $\theta = 5$, we found $\widehat{Bias}(e^{-\overline{X}}) = 0.000539$, and $\widehat{Var}(e^{-\overline{X}}) = 5 \times 10^{-6}$. For $n = 20$ and $\theta = 5$, we have $\widehat{Bias}(e^{-\overline{X}}) = 0.000879$, and $\widehat{Var}(e^{-\overline{X}}) = 1.6 \times 10^{-5}$.

It appears that as n decreases, our bias *and* variance increases.

2. Recall that the MISE of a estimator $\hat{f}$ was given by $\mathbb{E}||\hat{f} - f||_2^2$. In our case, we know the true distribution is Uniform on $(0,1)$ so $f(x) = 1$ for $x \in (0,1)$. This yields

$$MISE(\hat{f}) = \mathbb{E}\int_0^1 (\hat{f}(x) - 1)^2 dx = \mathbb{E}\int_0^1 \hat{f}(x)^2 dx - 2\mathbb{E}\int_0^1 \hat{f}(x)dx + \mathbb{E}\int_0^1 1dx$$

$$= \mathbb{E}\int_0^1 \hat{f}(x)^2 dx - 2 + 1$$

The integrand is positive in this case so we interchange the expectation and integral to achieve the following

$$= \int_0^1 \mathbb{E}\hat{f}(x)^2 dx - 1 = \int_0^1 \mathbb{V}(\hat{f}(x)) + \mathbb{E}(\hat{f}(x))^2 dx - 1 = \int_0^1 \mathbb{V}(\hat{f}(x))dx + \int_0^1 \big[\mathbb{E}(\hat{f}(x))\big]^2 dx - 1$$

Recall that $\mathbb{V}(\hat{f}) = \frac{1}{nh}\mathbb{V}(f_j)$ where f_j is the number of observations in bin B_j. Here, $f_j \sim \text{Binom}(n, \frac{1}{m})$ so $\frac{1}{(nh)^2}\mathbb{V}(f_j) = \frac{m^2}{n^2}\frac{n}{m}(1-\frac{1}{m}) = \frac{m-1}{n}$. Hence, we have

$$= \frac{m-1}{n} + \int_0^1 \left[\mathbb{E}(\hat{f}(x))\right]^2 dx - 1$$

Now recall that

$$\hat{f}(x) = \sum_{j=1}^{m} \frac{f_j}{nh} I(x \in B_j) \frac{m}{n} \sum_{j=1}^{m} f_j I(x \in B_j)$$

Now using the distribution of f_j we have

$$= \mathbb{E}(\hat{f}(x)) = \frac{m}{n}\sum_{j=1}^{m} \mathbb{E}(f_j) I(x \in B_j) = \sum_{j}^{m} I(x \in B_j)$$

Notice that for any x fixed, $x \in B_j$ for only one j. So $\mathbb{E}(\hat{f}(x)) = \sum_j^m I(x \in B_j) = 1$. Using this information, we now have

$$MISE(\hat{f}) = \frac{m-1}{n} + \int_0^1 (1)^2 dx - 1 = \frac{m-1}{n}$$

Now, it's clear that $MISE$ is minimized at $m = 1$. This follows intuition. If the true underlying distribution is uniform on $(0,1)$, then the histogram that best fits the data is one that only has one bin. That is it has a single bin over the support of X. In this way, we ensure that our histogram estimates the same density value for every point.

3. Let $X_1, \dots, X_n \overset{iid}{\sim} f(x)$ where $f(x) = \sum_{j=1}^{k} c_j I(x \in B_j)$. Note that $f(X) \in \{c_1, c_2, \dots, c_j\}$ so letting f_j be the number of X_k in bin B_j we can write the likelihood function.

$$\mathcal{L}(X_1, \dots, X_N | f) = f(X_1, X_2, \dots, X_n) = \prod_{j=1}^{k} f(X_j) = \prod_{j=1}^{k} c_j^{f_j}$$

While leads to the log-likelihood function

$$\ell(X_1, X_2, \dots, X_n) = \sum_{j=1}^{j} f_j \log(c_j)$$

Now, to maximize this function with respect to c_j, we would simply let each c_j grow without bound. Clearly this does not result in a density. In order for the c_j to give a density, we require $\sum_{j=1}^{k} hc_j = 1$ where h is the binwidth. Clearly, this requires that $\sum_{j=1}^{k} c_k = \frac{1}{h}$. Hence we look to maximize ℓ with respect to $\sum_{j=1}^{k} c_k = \frac{1}{h}$.

To solve this problem we will use the method of Lagrange multipliers. That is we will solve the system represented by $\nabla(f_1 \log(c_1), \dots, f_k \log(c_k)) = \lambda \nabla(c_1, \dots, c_k)$ and $\sum_{j=1}^{k} c_j = \frac{1}{h}$. Notice that the first set of equations is just given by $\frac{f_j}{c_j} = \lambda$ which implies $c_j = \frac{f_j}{\lambda}$. Plugging this into the final equation we have $\sum_{j=1}^{k} \frac{f_j}{\lambda} = \frac{n}{\lambda} = \frac{1}{h}$ which implies $\lambda = nh$. Hence, $\widehat{c_j}_{MLE} = \frac{f_j}{nh}$.

4. Recall $MISE(\hat{f}) \approx \frac{1}{h} + \frac{h^2}{12}\int(f'(x))^2 dx$. For ease of notation, let $I = \int(f'(x))^2 dx$. Using this, and the optimal bandwidth is given by $h_* = \left(\frac{6}{nI}\right)^{1/3}$. From here we can calculate the ratio of the following two MISE.

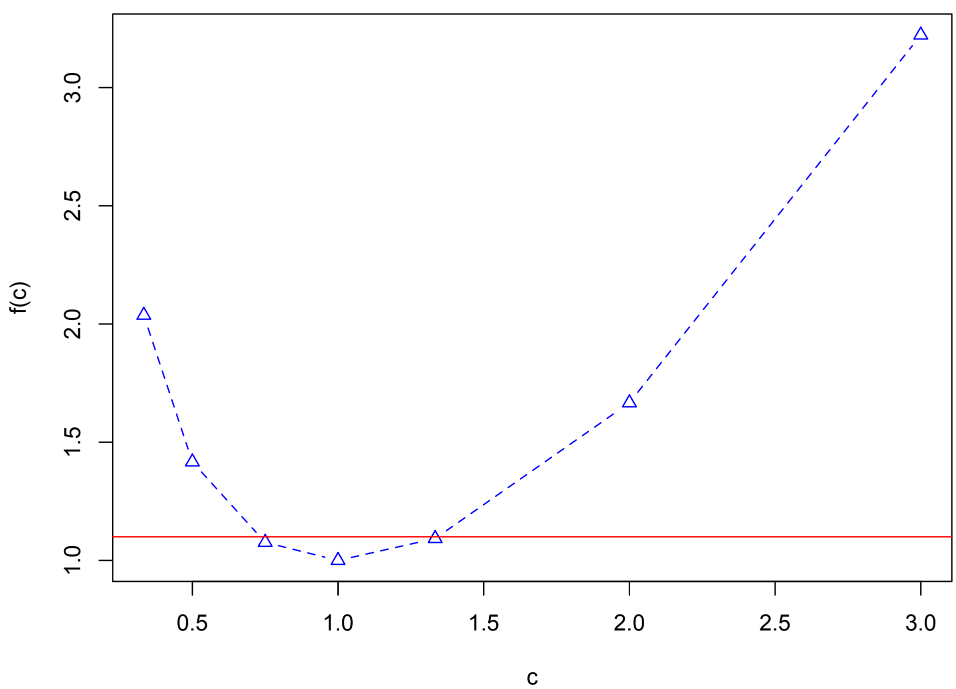
$$\frac{MISE(ch_*)}{MISE(h_*)} = \frac{\frac{1}{nch_*} + \frac{c^2h_*^2}{12}I}{\frac{1}{nh_*} + \frac{h_*^2}{12}I} = \frac{12 + Inc^3h_*^3}{12nch_*} \bigg/ \frac{12 + Inh_*^3}{12nh_*} = \frac{12nh_*\left(12 + Inc^3h_*^3\right)}{12nch_*\left(12 + Inh_*^3\right)}$$

$$= \frac{12 + Inc^3h_*^3}{c\left(12 + Inh_*^3\right)}$$

Now recall that $h_* = \left(\frac{6}{nI}\right)^{1/3}$ so $h_*^3 = \frac{6}{In}$. This gives

$$= \frac{12 + Inc^3\left(6/In\right)}{c\left(12 + In\left(6/In\right)\right)} = \frac{12 + 6c^3}{c(12+6)} = \frac{2 + c^3}{3c}$$

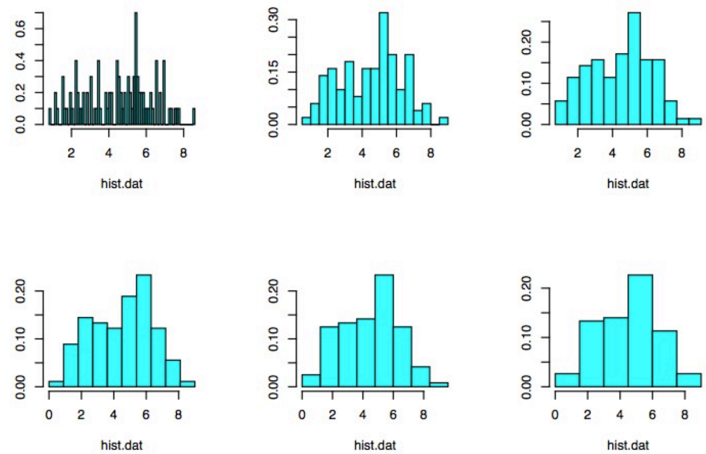
Now evaluating this proportional change at multiple values, c, we can plot the relationship at follows.



c	1/3	1/2	3/4	1	4/3	2	3
$f(x)$	2.04	1.42	1.08	1.00	1.09	1.67	3.22

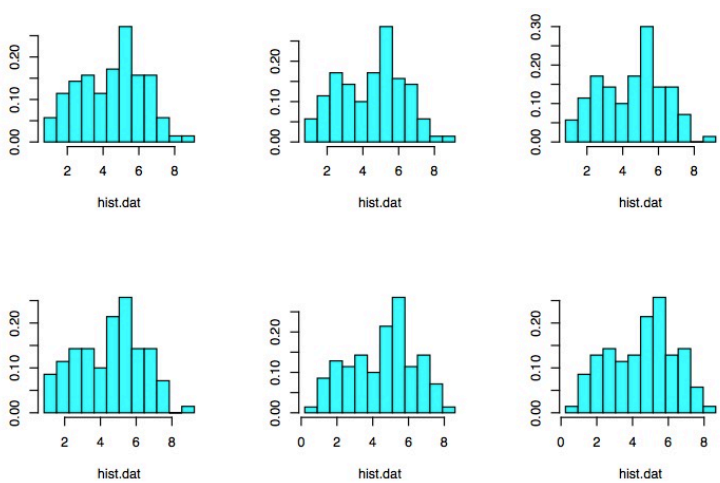
Here, the points, $c = 3/4, 1, 4/3$ result in changes less that 10% corresponding to $f(c) \leq 1.1$ depicted by the horizontal red line. Notice here that $f(c)$ grows rapidly for c small and gradually for c large. That is, the rate of change in the $MISE$ for c small is greater than the rate of change of the $MISE$ for c large. Hence, the $MISE$ of the histogram estimator is more sensitive to high variance than to high bias.

5. (a) Consider the histograms given below with varying bandwidths.



Notice that conclusions about the underlying distribution change as a function of the bandwidth. For h small (first two plots) there appears to be two, maybe three high density regions. In the third and fourth plot, it seems there are two. In the final two plots, there appears to be a single skewed distributions. Now, since we know that this data was generated from a mixture of two normal distributions centered at 3 and 6, we expect the data to appear as it does in third and fourth plots. The third plot better highlights the gap between the two distributions, so I choose this bandwidth corresponding to $h = 0.7$.

(b) Consider the histograms given below with varying "starting" points x_0.



The first, fifth, and sixth plot fail to highlight the two underlying distributions. In these plots, while visible, the data appears to come from a normal, possibly mixed with a uniform. In the second, third, and fourth plot, the starting positions do a nice job of separating the two normal distributions. Specifically, for plot 3, we see two clear normal distributions centered at 3 and 6. Thus I choose this starting point $x_0 = 0.1$.

It should be noted that we *knew* the distribution before choosing the bandwidth and starting parameters. If we hadn't know these parameters, these decisions would be much more difficult to make (if not entirely arbitrary).

MA 750: HW1

Benjamin Draves

September 26, 2017

Exercise 1

(d)

```
#set seed
set.seed(1665)

#set up tunning variables
n = 100
num_samp = 1000
mu = 5

#Get num_samp estimates of exp(-xbar) where -xbar from Pois(mu) with n samples
sim = replicate(num_samp, exp(-mean(rpois(n,mu))))

#compute empircal bias for each point
estimated_bias = sim - exp(-mu)

#find the mean bias
mean_bias = mean(estimated_bias)

#display estimated bias
paste("The estiamted bias of the MLE estimator:", round(mean_bias, 6))
```

```
## [1] "The estiamted bias of the MLE estimator: 0.000218"
```

```
#estimate the variance of the estimator
estimated_var = var(sim)

#display estimated variance
paste("The estiamted variance of the MLE estimator:", round(estimated_var, 6))
```

```
## [1] "The estiamted variance of the MLE estimator: 2e-06"
```

Compare estimates to the part approximated in part b

(e)

```
#set up tunning variables
n = 50
num_samp = 1000
mu = 5

#Get num_samp estimates of exp(-xbar) where -xbar from Pois(mu) with n samples
sim = replicate(num_samp, exp(-mean(rpois(n,mu))))

#compute empircal bias for each point
estimated_bias = sim - exp(-mu)
```

```
#find the mean bias
mean_bias = mean(estimated_bias)

#display estimated bias
paste("The estiamted bias of the MLE estimator:", round(mean_bias, 6))
```

```
## [1] "The estiamted bias of the MLE estimator: 0.000539"
```

```
#estimate the variance of the estimator
estimated_var = var(sim)

#display estimated variance
paste("The estiamted variance of the MLE estimator:", round(estimated_var, 6))
```

```
## [1] "The estiamted variance of the MLE estimator: 5e-06"
```

Compare estimates here

```
#set up tunning variables
n = 20
num_samp = 1000
mu = 5

#Get num_samp estimates of exp(-xbar) where -xbar from Pois(mu) with n samples
sim = replicate(num_samp, exp(-mean(rpois(n,mu))))

#compute empircal bias for each point
estimated_bias = sim - exp(-mu)

#find the mean bias
mean_bias = mean(estimated_bias)

#display estimated bias
paste("The estiamted bias of the MLE estimator:", round(mean_bias, 6))
```

```
## [1] "The estiamted bias of the MLE estimator: 0.000879"
```

```
#estimate the variance of the estimator
estimated_var = var(sim)

#display estimated variance
paste("The estiamted variance of the MLE estimator:", round(estimated_var, 6))
```

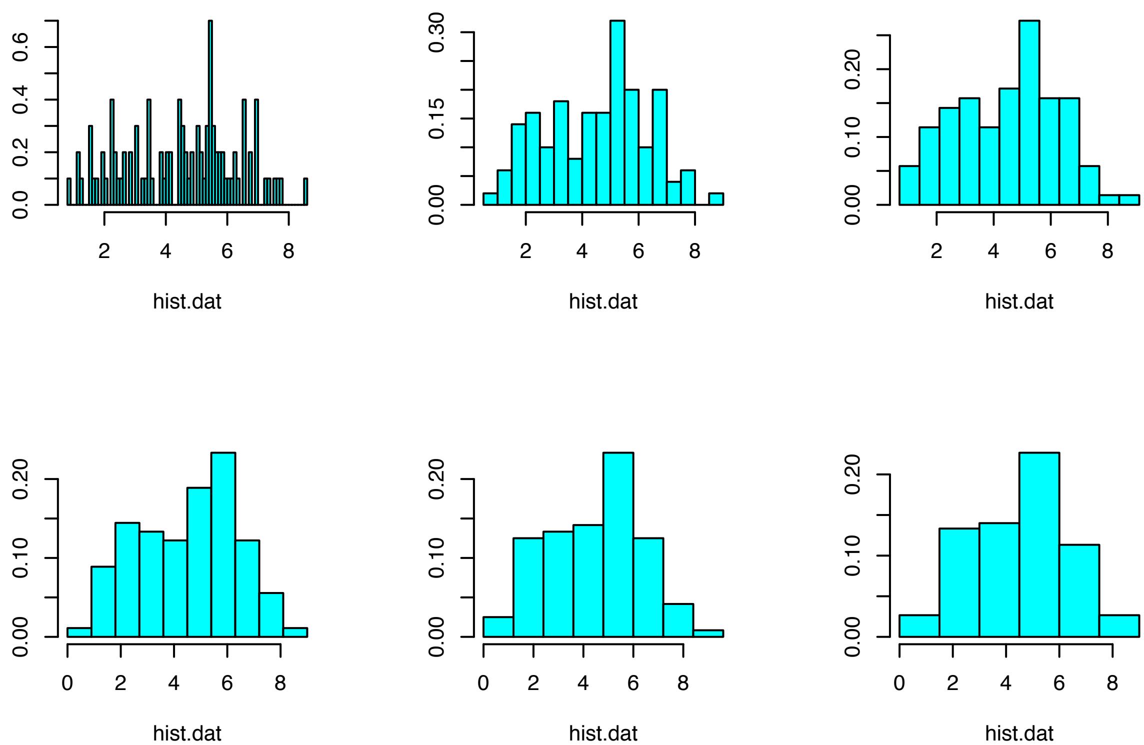
```
## [1] "The estiamted variance of the MLE estimator: 1.6e-05"
```

Compare estimates here

Exericse 5

```
#get sample
dat = matrix(0,nrow = 100, ncol = 2)
dat[,1] = rnorm(100, mean = 3, sd = 1)
dat[,2] = rnorm(100, mean = 6, sd = 1)

#get indicator
mixture1 = rbinom(n = 100, size = 1, p = .4)
```

```
indicator = matrix(0, nrow = 100, ncol = 2)
indicator[,1] = mixture1
indicator[,2] = ifelse(indicator[,1] == 0, 1, 0)

#get histogram data
hist.dat = rowSums(dat*indicator)

#load necessary packages
library(MASS)

#Check out histograms
par(mfrow = c(2.,3))
truehist(hist.dat, h = 0.1, x0 = 0)
truehist(hist.dat, h = 0.5, x0 = 0)
truehist(hist.dat, h = 0.7, x0 = 0)
truehist(hist.dat, h = 0.9, x0 = 0)
truehist(hist.dat, h = 1.2, x0 = 0)
truehist(hist.dat, h = 1.5, x0 = 0)
```




```
#Check out histograms
par(mfrow = c(2.,3))
truehist(hist.dat, h = 0.7, x0 = 0)
truehist(hist.dat, h = 0.7, x0 = 0.05)
truehist(hist.dat, h = 0.7, x0 = 0.1)
truehist(hist.dat, h = 0.7, x0 = 0.15)
truehist(hist.dat, h = 0.7, x0 = 0.2)
truehist(hist.dat, h = 0.7, x0 = 0.25)
```