Fitting a density curve to a histogram in R

Is there a function in R that fits a curve to a histogram?

Let's say you had the following histogram

```
hist(c(rep(65, times=5), rep(25, times=5), rep(35, times=10), rep(45, times=4)))
```

It looks normal, but it's skewed. I want to fit a normal curve that is skewed to wrap around this histogram.

This question is rather basic, but I can't seem to find the answer for R on the internet.

r   histogram   curve-fitting   r-faq

edited Nov 16 '12 at 16:38   asked Sep 30 '09 at 11:23

Gregor
23.9k   4   35   77

user5243421
1,773   8   46   82

Do you want to find m and s such that the Gaussian distribution N(m,s) fits to your data? – norheim.se Sep 30 '09 at 11:38

I'm not sure what that means... >_> –   user5243421   Sep 30 '09 at 11:41

9   @mathee: I think he means m = mean, and s = standard deviation. Gaussian distribution is another name for normal distribution. – Peter Mortensen Sep 30 '09 at 11:54

+1 for politeness. –   user5243421   Dec 6 '09 at 12:01

5 Answers

If I understand your question correctly, then you probably want a density estimate along with the histogram:

```r
X <- c(rep(65, times=5), rep(25, times=5), rep(35, times=10), rep(45, times=4))
hist(X, prob=TRUE)              # prob=TRUE for probabilities not counts
lines(density(X))              # add a density estimate with defaults
lines(density(X, adjust=2), lty="dotted")   # add another "smoother" density
```

Edit a long while later:

Here is a slightly more dressed-up version:

```r
X <- c(rep(65, times=5), rep(25, times=5), rep(35, times=10), rep(45, times=4))
hist(X, prob=TRUE, col="grey")# prob=TRUE for probabilities not counts
lines(density(X), col="blue", lwd=2) # add a density estimate with defaults
lines(density(X, adjust=2), lty="dotted", col="darkgreen", lwd=2)
```

along with the graph it produces:

Histogram of X
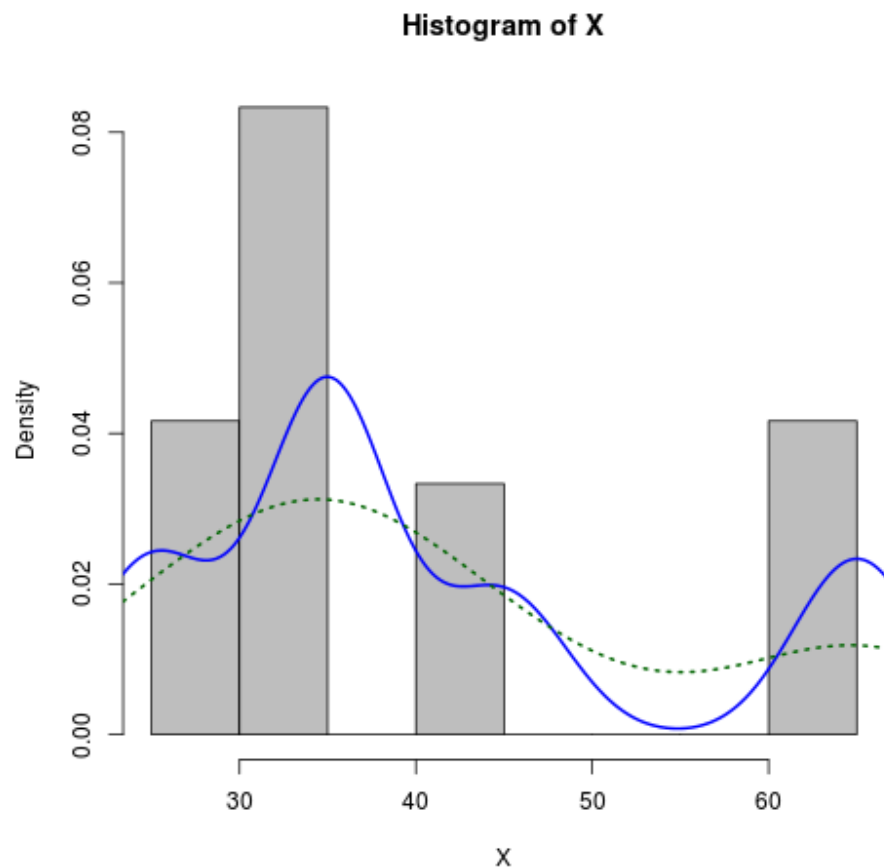
answered Sep 30 '09 at 12:02

Dirk Eddelbuettel
192k   22   344   442

---

3   Another success by Dirk Eddelbuettel! Cheers – Federico Giorgi Jun 25 '13 at 18:58

---

1   +1 - can you also do it the other way around, i.e. adjusting the density plot to fit the histogram? – vonjd Nov 14 '13 at 10:20

---

1   I suggest giving additional parameter to `lines(density(X,na.rm= TRUE)` as the vector may contain NA values. – Anirudh Jan 26 '14 at 4:56
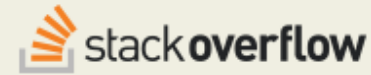
Such thing is easy with ggplot2

```
library(ggplot2)
dataset <- data.frame(X = c(rep(65, times=5), rep(25, times=5), rep(35, times=10), rep(45,
times=4)))
ggplot(dataset, aes(x = X)) + geom_histogram(aes(y = ..density..)) + geom_density()
```

or to mimic the result from Dirk's solution

```
ggplot(dataset, aes(x = X)) + geom_histogram(aes(y = ..density..), binwidth = 5) +
geom_density()
```

answered Sep 30 '09 at 18:30

Thierry
10.1k   2   18   48

Here's the way I do it:

```
foo <- rnorm(100, mean=1, sd=2)
hist(foo, prob=TRUE)
curve(dnorm(x, mean=mean(foo), sd=sd(foo)), add=TRUE)
```

A bonus exercise is to do this with ggplot2 package ...

edited May 4 '12 at 12:06   answered Sep 30 '09 at 13:32

Mike T   John Johnson
13.4k   4   41   73   **265**   1   5

However, if you want something that is skewed, you can either do the density example from above,
transform your data (e.g. foo.log <- log(foo) and try the above), or try fitting a skewed distribution, such
as the gamma or lognormal (lognormal is equivalent to taking the log and fitting a normal, btw). –
John Johnson Sep 30 '09 at 13:35

2   But that still requires estimating the parameters of your distribution first. – Dirk Eddelbuettel Sep 30 '09 at 13:48

This gets a bit far afield from simply discussing R, as we are getting more into theoretical statistics, but you might try this link for the Gamma: en.wikipedia.org/wiki/Gamma_distribution#Parameter_estimation For lognormal, just take the log (assuming all data is positive) and work with log-transformed data. For anything fancier, I think you would have to work with a statistics textbook. – John Johnson Sep 30 '09 at 14:45

1   I think you misunderstand how both the original poster as well as all other answers are quite content to use non-parametric estimates -- like an old-school histogram or a somewhat more modern data-driven densisty estimate. Parametric estimates are great if you have good reason to suspect a distribution. But that was not the case here. – Dirk Eddelbuettel Sep 30 '09 at 19:25

Dirk has explained how to plot the density function over the histogram. But sometimes you might want to go with the stronger assumption of a skewed normal distribution and plot that instead of density. You can estimate the parameters of the distribution and plot it using the sn package:

```
> sn.mle(y=c(rep(65, times=5), rep(25, times=5), rep(35, times=10), rep(45, times=4)))
$call
sn.mle(y = c(rep(65, times = 5), rep(25, times = 5), rep(35,
    times = 10), rep(45, times = 4)))

$cp
    mean      s.d. skewness
41.46228 12.47892  0.99527
```
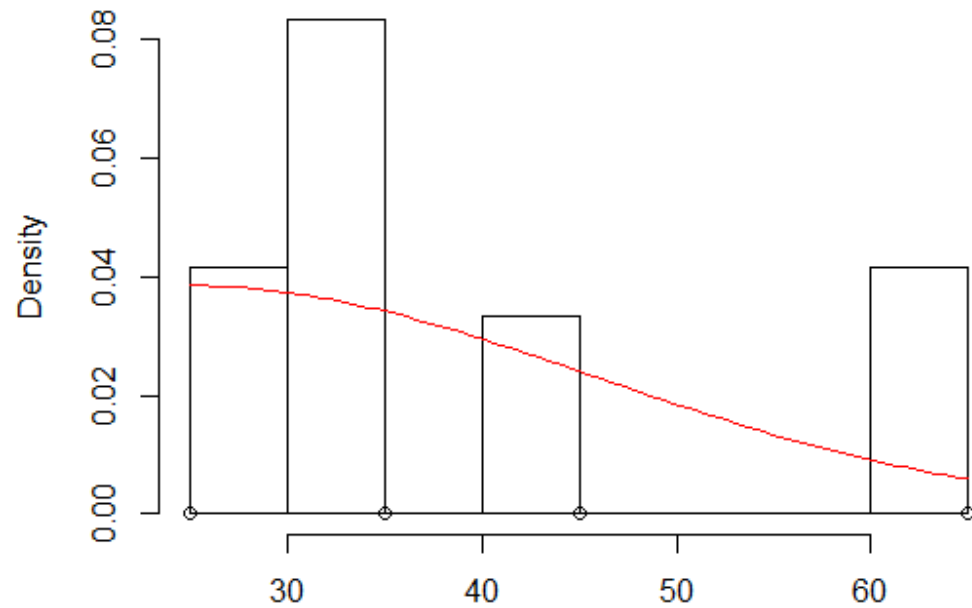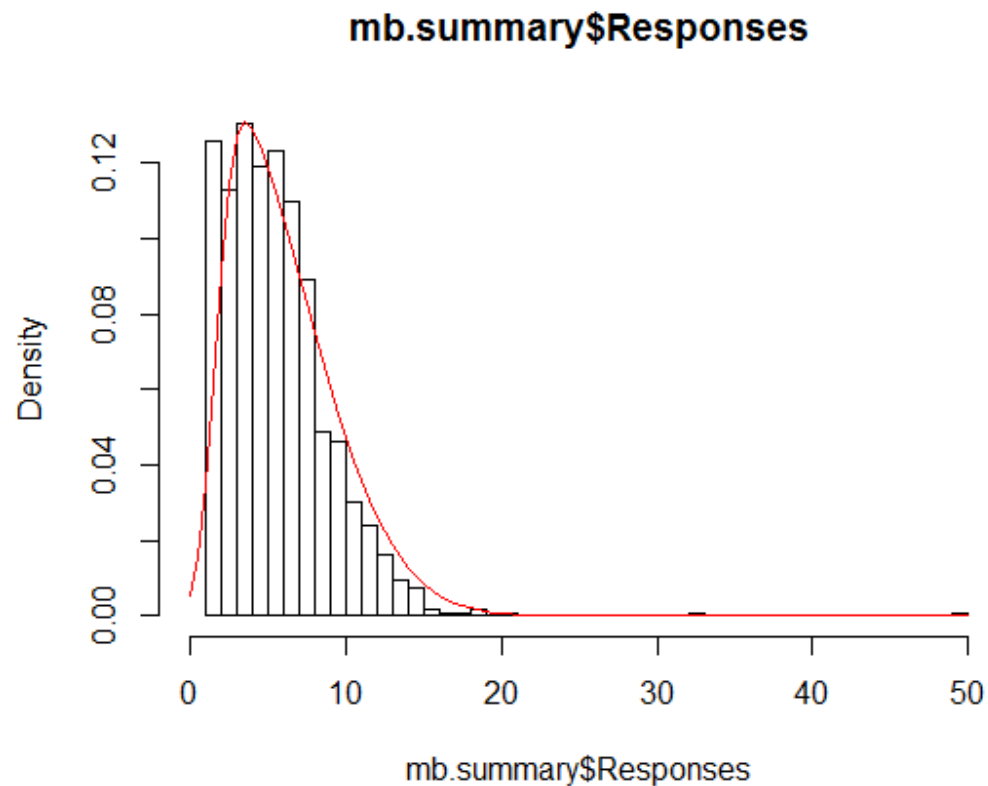
c(rep(65, times = 5), rep(25, times = 5), rep(35, times = 10),
rep(45, times = 4))

This probably works better on data that is more skew-normal:

mb.summary$Responses

I had the same problem but Dirk's solution didn't seem to work. I was getting this warning messege every time

```
"prob" is not a graphical parameter
```

I read through ?hist and found about `freq` : a logical vector set TRUE by default.

the code that worked for me is

```
hist(x,freq=FALSE)
lines(density(x),na.rm=TRUE)
```

answered Jan 21 '14 at 14:34

Matias Andina

715   1   8   19