

Homework # 1

1. Since we have $X \text{ HypG}(12, \theta, 5)$, we can use the function `dhyper()` to obtain $Pr(X = x|\theta)$ for specific values of x and θ .
2. To draw one gold ball for a specific value of θ , have the following function

```
hyp_theta <- function(theta, x = 1){  
  dhyper(x, theta, 12 - theta, 5)  
}
```

3. To create a 2-way table of $Pr(X = x|\theta)$ for $X = 0, \dots, 5$ and $\theta = 0, \dots, 12$ using the function above, we can use the `mapply()` function by the following.

```
matrix(mapply(hyp_theta, theta = c(rep(0, 6), rep(1,6), rep(2,6), rep(3,6),  
  rep(4,6), rep(5, 6), rep(6, 6), rep(7,6), rep(8,6), rep(9,6), rep(10,6),  
  rep(11, 6), rep(12,6)), x = c(rep(0:5, 13))), 13, 6, byrow = T)
```

Another option in building this table is to use a `for` loop as in the code below:

```
tab1 <- matrix(rep(NA, 6*13), 13, 6)  
for(i in 0:5){  
  tab1[, i+1] <- dhyper(i, 0:12, 12:0,5)  
}
```

Code is provided in the appendix to clean this up into the table given below.

	0	1	2	3	4	5	Row Sums
0	1.000	0.000	0.000	0.000	0.000	0.000	1
1	0.583	0.417	0.000	0.000	0.000	0.000	1
2	0.318	0.530	0.152	0.000	0.000	0.000	1
3	0.159	0.477	0.318	0.045	0.000	0.000	1
4	0.071	0.354	0.424	0.141	0.010	0.000	1
5	0.027	0.221	0.442	0.265	0.044	0.001	1
6	0.008	0.114	0.379	0.379	0.114	0.008	1
7	0.001	0.044	0.265	0.442	0.221	0.027	1
8	0.000	0.010	0.141	0.424	0.354	0.071	1
9	0.000	0.000	0.045	0.318	0.477	0.159	1
10	0.000	0.000	0.000	0.152	0.530	0.318	1
11	0.000	0.000	0.000	0.000	0.417	0.583	1
12	0.000	0.000	0.000	0.000	0.000	1.000	1
Col Sums	2.167	2.167	2.167	2.167	2.167	2.167	

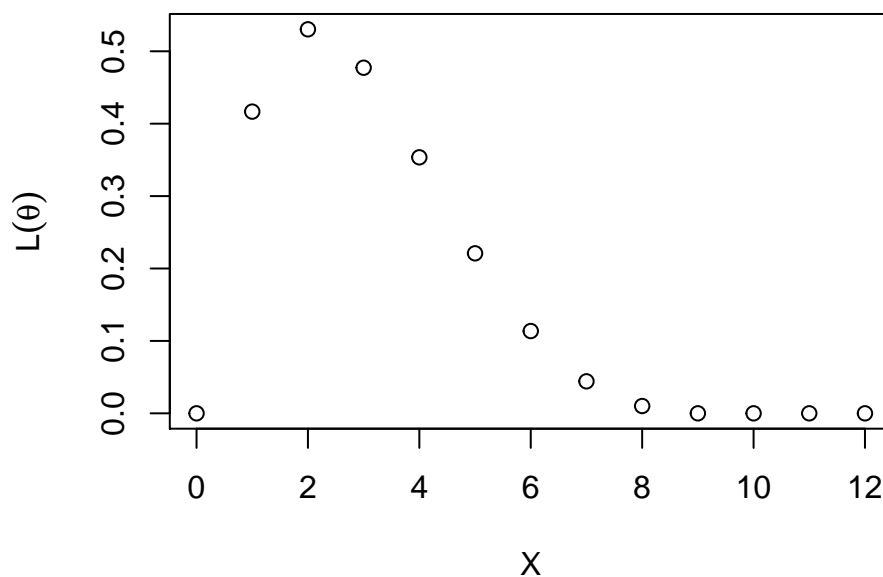
The columns are for specific values of X and the rows are for specific values of θ .

4. (combined in 3 and 5)
5. In the table above, notice that the rows sum to 1. This is because $Pr(X = x|\theta)$ is a probability function across its entire domain. However, the columns do not sum to 1. This shows that the columns of the table do not represent a probability function. Rather, it is simply a function of θ given specific values of θ . For this reason, we do not use $Pr(\theta|X)$ for our notation, but rather we use $L(\theta|X)$. We use L for "likelihood" to describe values of θ given an outcome X .
6. To find our "Maximum Likelihood Estimator", we find the max of $L(\theta|X)$. In this context, a discrete one, we simply take the largest of the values in each column. For instance, if we observe 3 gold balls in our sample of 5, our MLE for θ is $\hat{\theta} = 7$ since the .442 is the largest number in the column associated with $X = 3$.
7. An **estimator** is a process that yields an **estimate** given the data. Thus, when we talk about an **estimator**, we are talking about the process that yields different **estimates** given the data. Whereas, when we say **estimate**, we have observed the data and can now calculate the **estimate** from our **estimator**.
8. The Maximum Likelihood **Estimator** for this problem can take on the values $\{0, 2, 5, 7, 10, 12\}$. We can obtain these values by finding the largest value in the table above for each column, where the row for the maximum value gives the associated Maximum Likelihood **Estimate** given the value of X , the observed data.

9. The table below gives the maximum likelihood estimates associated with the different possible values of X . For each observed X , we have an associated value of $\hat{\theta}$

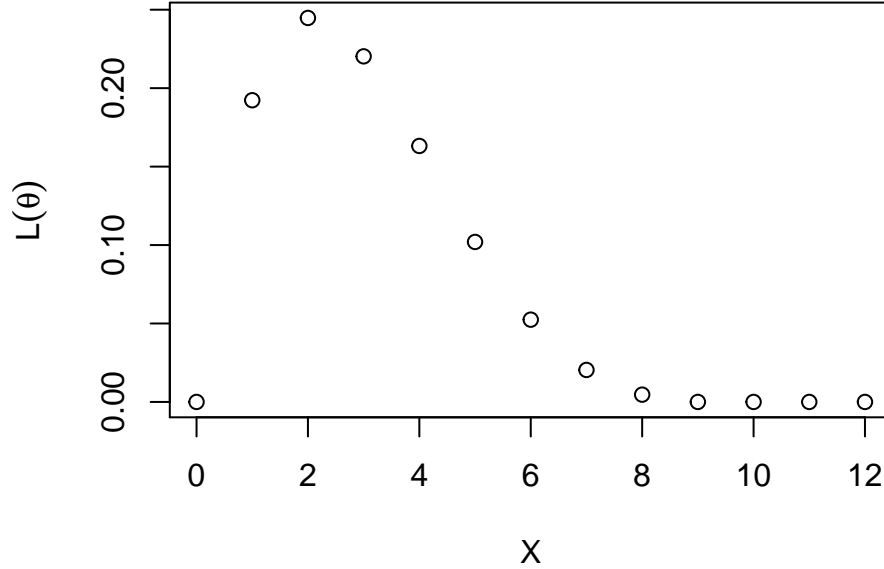
X	$\hat{\theta}$
0	0
1	2
2	5
3	7
4	10
5	12

10. (a) If we observe 1 gold ball, we have $\hat{\theta} = 2$. That is, if in our experiment, one of the five balls drawn is gold, our estimate of the true number of gold balls in the bucket based on the maximum likelihood estimator is 2.
- (b) Given that we observe 1 gold ball, we have the following plot of the likelihood function.



We do not plot a histogram here as that would suggest that values between the integers from 0 to 12 are meaningful. However, in the context of this problem, the number of balls in the bucket can only take on integer values (from 0 to 12).

(c) Below, we plot the "normalized" likelihood function.



They are "normalized" in that now $L(\theta|X = 1)$ represents a probability function in that summing over all possible values of θ gives 1 rather than 2.166 as in before.

11. (a) When we use the function `lm()`, R is finding the Best Linear Unbiased Estimators (BLUE's) for β by calculating $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ where we the model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ with $\epsilon \stackrel{iid}{\sim} (0, \sigma^2 \mathbf{I})$. Whereas, when we use the function `glm()`, R is using an iteratively reweighted (IWLS) algorithm to calculate the Maximum Likelihood Estimates for β (and σ^2). Here we still have the model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ with $\epsilon \stackrel{iid}{\sim} (0, \sigma^2 \mathbf{I})$. Now, we no longer have BLUE's, however the MLE's are asymptotically efficient (as sample size goes to infinity).
- (b) Confidence intervals for each are obtained in quite a different manner. Consider the `confint()` function in R which has methods for both `lm()` and `glm()`. For `lm()`, R simply calculates the confidence interval as $\hat{\beta} \pm t_{n-k, \alpha} \times SE(\hat{\beta})$ where k is the number of parameters in the model. However, in the method for `glm()`, R calculates confidence intervals through profiling the associated likelihoods. In the details section for this function (from the MASS package), the authors describe the method for making this calculation. They write, "These confint methods call the appropriate profile method, then find the confidence intervals by interpolation in the profile traces."

12. We have

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

In the context of the bucket problem, we have

$$\begin{aligned}
P(\theta|x) &= \frac{P(x|\theta)P(\theta)}{P(x)} \\
&= \frac{f(x|\theta)g(\theta)}{h(x)} \\
&= \frac{\binom{\theta}{x} \binom{12-\theta}{5-x}}{\binom{12}{5}} g(\theta) \\
&= \frac{f(x|\theta)}{h(x)} g(\theta)
\end{aligned}$$

However, we need to know (or assume) $g(\theta)$, our prior distribution, to finish the calculation.

13. As stated in the previous problem, to obtain the posterior distribution for the bucket problem, we must first assume a prior distribution. We shall, as before, call this $g(\theta)$. Thus, we can use the likelihood function to calculate the posterior distribution as follows:

$$\begin{aligned}
P(\theta|x) &= \frac{P(x|\theta)P(\theta)}{P(x)} \\
&= \frac{f(x|\theta)g(\theta)}{h(x)} \\
&= \frac{L(\theta|X)g(\theta)}{h(x)} \\
&= \frac{L(\theta|X)g(\theta)}{\sum_{i=0}^{12} L(\theta_i|X)g(\theta_i)}
\end{aligned}$$

14. (a) For this prior, we have

θ	$g(\theta)$	$L(\theta X = 1)$	$L(\theta X = 1) * g(\theta)$	$f(\theta X = 1)$
0	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.4167	0.0000	0.0000
2	0.0278	0.5303	0.0147	0.1230
3	0.0556	0.4773	0.0265	0.2213
4	0.0833	0.3535	0.0295	0.2459
5	0.1111	0.2210	0.0246	0.2049
6	0.1389	0.1136	0.0158	0.1317
7	0.1667	0.0442	0.0074	0.0615
8	0.1389	0.0101	0.0014	0.0117
9	0.1111	0.0000	0.0000	0.0000
10	0.0833	0.0000	0.0000	0.0000
11	0.0556	0.0000	0.0000	0.0000
12	0.0278	0.0000	0.0000	0.0000

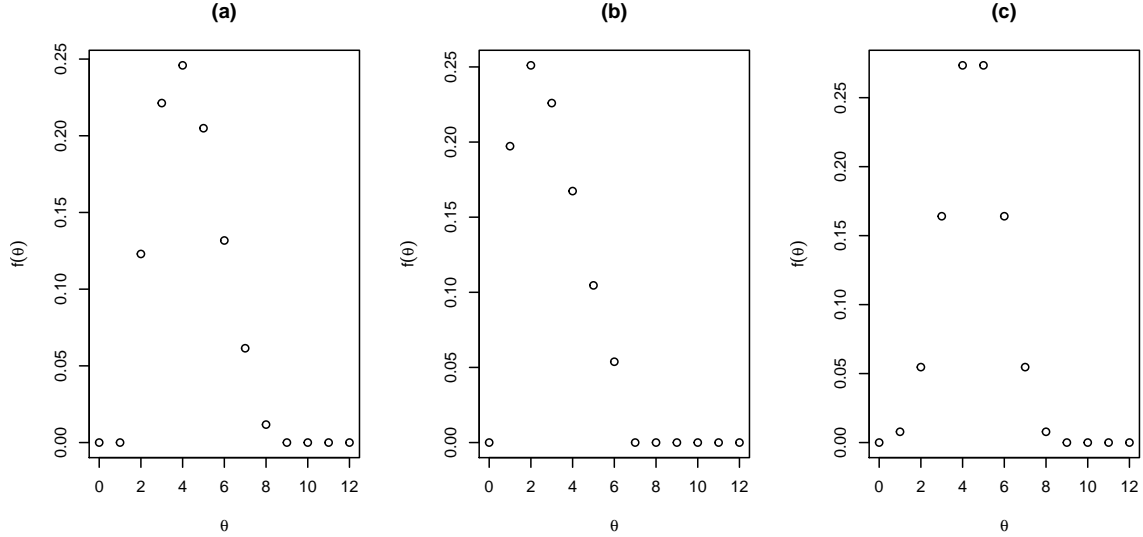
(b) For this prior, we have

θ	$g(\theta)$	$L(\theta X = 1)$	$L(\theta X = 1) * g(\theta)$	$f(\theta X = 1)$
0	0.0000	0.0000	0.0000	0.0000
1	0.1667	0.4167	0.0694	0.1973
2	0.1667	0.5303	0.0884	0.2510
3	0.1667	0.4773	0.0795	0.2259
4	0.1667	0.3535	0.0589	0.1674
5	0.1667	0.2210	0.0368	0.1046
6	0.1667	0.1136	0.0189	0.0538
7	0.0000	0.0442	0.0000	0.0000
8	0.0000	0.0101	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0000
11	0.0000	0.0000	0.0000	0.0000
12	0.0000	0.0000	0.0000	0.0000

(c) For this prior, we have

θ	$g(\theta)$	$L(\theta X = 1)$	$L(\theta X = 1) * g(\theta)$	$f(\theta X = 1)$
0	0.0002	0.0000	0.0000	0.0000
1	0.0029	0.4167	0.0012	0.0078
2	0.0161	0.5303	0.0085	0.0547
3	0.0537	0.4773	0.0256	0.1641
4	0.1208	0.3535	0.0427	0.2734
5	0.1934	0.2210	0.0427	0.2734
6	0.2256	0.1136	0.0256	0.1641
7	0.1934	0.0442	0.0085	0.0547
8	0.1208	0.0101	0.0012	0.0078
9	0.0537	0.0000	0.0000	0.0000
10	0.0161	0.0000	0.0000	0.0000
11	0.0029	0.0000	0.0000	0.0000
12	0.0002	0.0000	0.0000	0.0000

15. We have the following plots of $f(\theta|X = 1)$ for the three preceeding situations.



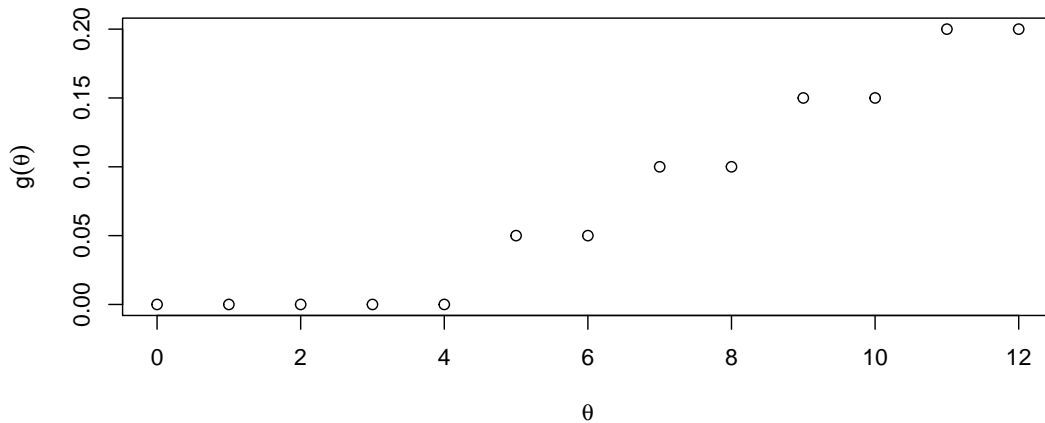
The probabilities shown in the plots make intuitive sense for the three situations given each prior. For (a), we have $f(\theta|X = 1) = 0$ for $\theta \in \{0, 1, 9, 10, 11, 12\}$ since $g(\theta) = 0$ for $\theta \in \{0, 1, 3\}$ and $L(\theta|X = 1) = 0$ for $\theta \in \{9, 10, 11, 12\}$. In fact, since all three share the same likelihood, $f(\theta|X = 1) = 0$ for $\theta \in \{, 10, 11, 12\}$ in each case. In (b), we see that $f(\theta|X = 1) = 0$ additionally for $\{7, 8\}$ since the prior probabilities are zero for each. In (c), the only values for which $f(\theta|X = 1) = 0$ are the ones for which $L(\theta|X = 1) = 0$ since all prior probabilities are positive.

16. None of the posterior distributions above match the normalized likelihood function in problem 10 since the process of normalizing the likelihood function is equivalent to

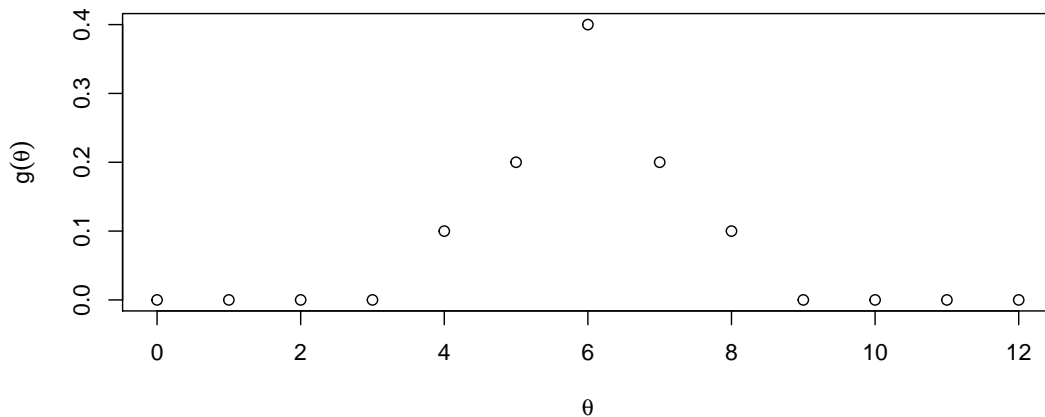
picking the non-informative prior. That is, we have $g(\theta) = \frac{1}{13} I(\theta)$, putting equal probability on each point of the prior.

17. The situation is complicated in that we have to make some somewhat non-informed decisions about our priors. For instance in (a), since we know the filler has a strong preference for the color blue, we have to decide what exactly that means. Do we think that the probability of the prior for say $\theta = 4$ is much higher than for $\theta = 8$, or do we think $P(\theta = 8) = 0$? Below, we draw some possible shapes for the prior distributions given the problem descriptions.

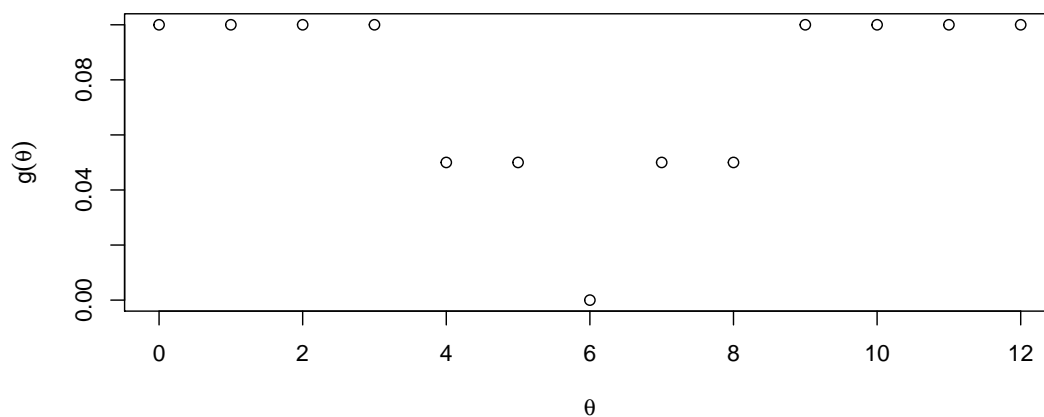
- (a) For a "strong" preference for blue, a possible prior distribution we might assume is



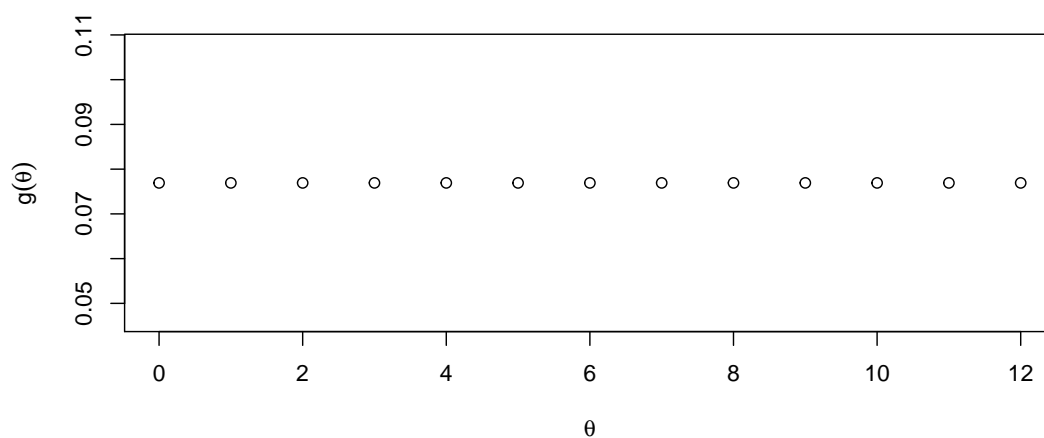
- (b) If we think the fillers almost always chose equal composition of gold and blue, we may have a prior that looks like this.



- (c) If we notice that all fillers put more of one color than the other, we may assume a prior that looks like this.



- (d) If we have no knowledge about how they might fill the bucket, we may just want to put equal probabilities on each outcome.



R-Code

```
## Require Packages
require(xtable)

# Problem 3
tab1 <- matrix(rep(NA, 6*13), 13, 6)
for(i in 0:5){
  tab1[, i+1] <- dhyper(i, 0:12, 12:0,5)
}

tab1 <- cbind(tab1, apply(tab1, 1, sum))
tab1 <- rbind(tab1, apply(tab1, 2, sum))

dimnames(tab1) <- list(c(0:12, "Col Sums"), c(0:5, "Row Sums"))
tab1 <- as.data.frame(tab1)
tab1[14,7] <- ""
xtable(tab1, digits = 3)

# Problem 9
plot(0:12, tab1[1:13,2], xlab = "X", ylab = expression(L(theta)))

#Problem 14
#a
g <- c(0,0,1/36,2/36,3/36,4/36,5/36,6/36,5/36,4/36,3/36,2/36,1/36)
l <- tab1[1:13,2]
gl <- g*l
f <- gl / sum(gl)
tab2 <- cbind(0:12, g, l, gl, f)
xtable(tab2, digits = 4)

#b
g <- c(0,1/6,1/6,1/6,1/6,1/6,1/6,0,0,0,0,0,0)
l <- tab1[1:13,2]
gl <- g*l
f <- gl / sum(gl)
tab3 <- cbind(0:12, g, l, gl, f)
xtable(tab3, digits = 4)

#c
g <- dbinom(0:12, 12, .5)
l <- tab1[1:13,2]
gl <- g*l
f <- gl / sum(gl)
```

```

tab4 <- cbind(0:12, g, l, gl, f)
xtable(tab4, digits = 4)

#Problem 17
#a
par(mfrow = c(1,1))
g <- c(0,0,0,0, 0,1/20, 1/20, 1/10, 1/10, 3/20, 3/20, 2/10, 2/10)
plot(0:12, g, xlab = expression(theta), ylab = expression(g(theta)))
#b
g <- c(0,0,0,0, 1/10,1/5, 2/5, 1/5, 1/10, 0, 0, 0, 0)
plot(0:12, g, xlab = expression(theta), ylab = expression(g(theta)))
#c
g <- c(1/10,1/10,1/10,1/10, 1/20,1/20, 0, 1/20, 1/20, 1/10, 1/10, 1/10, 1/10)
plot(0:12, g, xlab = expression(theta), ylab = expression(g(theta)))
#d
g <- rep(1/13, 13)
plot(0:12, g, xlab = expression(theta), ylab = expression(g(theta)))

```