Chase Joyner

882 Homework 2

September 28, 2016

Problem 1:

(a) The Jeffrey's prior distribution is obtained calculating the square root of the determinant of the fisher information matrix. Namely, for parameters $\boldsymbol{\theta}$, impose the prior distribution

$$\pi(\theta) \propto \sqrt{|\mathcal{I}(\boldsymbol{\theta})|}$$

where $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix and $|\cdot|$ is the determinant function. In a model with only one parameter, the Fisher information is just a scalar and so $|\mathcal{I}(\theta)| = \mathcal{I}(\theta)$. The Jeffrey's prior is invariant under reparameterization and therefore a new posterior under a different parameterization can easily be obtained without repeating the derivations. Another property of the Jeffrey's prior, though favorable or not is left for the statistician to decide, is that it acts as a non-informative prior. This could be beneficial when there is little to no knowledge of the parameter space. A drawback worth mentioning is that the computation of Jeffrey's prior can be tedious, depending on the situation. The complexity of calculating the Fisher information grows as the dimension of $\boldsymbol{\theta}$ gets larger because the Fisher information matrix is a $p \times p$ matrix, where p is the number of parameters. In fact, in some situations the Jeffrey's prior in high dimensions can lead to unfavorable results.

(b) Assume $Y_i \overset{iid}{\sim}$ Bernoulli(p). Then, the likelihood function is

$$f(p|\mathbf{Y}) = p^{\sum_{i=1}^{n} Y_i}(1-p)^{n-\sum_{i=1}^{n} Y_i}.$$

Taking the log, and denoting $Z = \sum_{k=1}^{n} Y_k$, we have the log likelihood function to be

$$\ell(p|\mathbf{Y}) = \log f(p|\mathbf{Y}) = Z \log p + (n - Z) \log(1 - p).$$

Now we can calculate the Fisher information. Using the fact that ℓ here is twice differentiable, we will use the alternate formulation, namely

$$\mathcal{I}(p) = -E\left[\frac{\partial^2}{\partial p^2} \ell(p|\mathbf{Y})\right] = -E\left[\frac{\partial}{\partial p}\left(\frac{Z}{p} - \frac{n-Z}{1-p}\right)\right]$$

$$= -E\left[-\frac{Z}{p^2} - \frac{n-Z}{(1-p)^2}\right] = \frac{1}{p^2}E[Z] + \frac{n-E[Z]}{(1-p)^2}.$$

Note that $Z = \sum_{k=1}^{n} Y_k \sim$ Binomial(n, p), and thus $E[Z] = np$. Therefore, we see that

$$\mathcal{I}(p) = \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p(1-p)}.$$

Finally, the Jeffrey's prior for this setting is

$$\pi(p) \propto \sqrt{|\mathcal{I}(p)|} \propto \frac{1}{\sqrt{p(1-p)}} = p^{\frac{1}{2}-1}(1-p)^{\frac{1}{2}-1}.$$

We see that Jeffrey's prior here is Beta$(1/2, 1/2)$. Now we can formulate the posterior distribution for p. Notice that

$$
\begin{aligned}
\pi(p|\mathbf{Y}) &\propto \pi(\mathbf{Y}|p)\pi(p) \\
&\propto p^{\sum_{k=1}^{n} Y_k}(1-p)^{n-\sum_{k=1}^{n} Y_k} \cdot p^{\frac{1}{2}-1}(1-p)^{\frac{1}{2}-1} \\
&= p^{\sum_{k=1}^{n} Y_k + \frac{1}{2} - 1}(1-p)^{n-\sum_{k=1}^{n} Y_k + \frac{1}{2} - 1}.
\end{aligned}
$$

Then, we see that imposing a Jeffrey's prior results in a Beta posterior, that is we have $p|\mathbf{Y} \sim \text{Beta}\left(n\overline{Y} + 1/2, n - n\overline{Y} + 1/2\right)$. Since the posterior distribution is a known distribution, to conduct posterior inference we can report any point estimates we want to, such as the posterior mean of p, the posterior mode of p, etc.. Also, we can calculate posterior $(1 - \alpha)\%$ quantiles very easily or use HPD intervals.

(c) Now we assume our data follows a Poisson model, i.e. $Y_i \overset{iid}{\sim} \text{Poisson}(\lambda)$. The log likelihood function is then

$$\ell(\lambda|\mathbf{Y}) = \sum_{k=1}^{n} Y_k \log \lambda - n\lambda + \sum_{k=1}^{n} \log(Y_k!).$$

Setting $Z = \sum_{k=1}^{n} Y_k$, we can calculate the Fisher information as

$$\mathcal{I}(\lambda) = -E\left[\frac{\partial^2}{\partial p^2} \ell(\lambda|\mathbf{Y})\right] = -E\left[\frac{Z}{\lambda} - \lambda\right] = -E\left[-\frac{Z}{\lambda^2}\right] = \frac{1}{\lambda^2} E[Z].$$

Notice that $Z = \sum_{k=1}^{n} Y_k \sim \text{Poisson}(n\lambda)$ and hence $E[Z] = n\lambda$. Then,

$$\mathcal{I}(\lambda) = \frac{1}{\lambda} n\lambda = \frac{n}{\lambda}.$$

Finally, the Jeffrey's prior in this situation is

$$\pi(\lambda) \propto \sqrt{|\mathcal{I}(\lambda)|} \propto \frac{1}{\sqrt{\lambda}}.$$

The posterior distribution is then

$$\pi(\lambda|\mathbf{Y}) \propto \lambda^{\sum_{k=1}^{n} Y_k} e^{-n\lambda} \cdot \frac{1}{\sqrt{\lambda}} = \lambda^{\sum_{k=1}^{n} Y_k + \frac{1}{2} - 1} e^{-n\lambda}.$$

We see that the posterior distribution is $\lambda|\mathbf{Y} \sim \text{Gamma}(\sum_{k=1}^{n} Y_k + 1/2, n)$. Here, we again can recognize the posterior distribution and so posterior inference is easy. We can report posterior point estimates by calculating the mean, median, mode, etc. of the posterior distribution as well as quantiles.

(d) For the last situation, assume $Y_i \overset{iid}{\sim} \text{Exp}(\beta)$. The log likelihood function is

$$\ell(\beta|\mathbf{Y}) = n \log \beta - \beta \sum_{k=1}^{n} Y_k.$$

Setting $Z = \sum_{k=1}^{n} Y_k$, the Fisher information is

$$\mathcal{I}(\beta) = -E\left[\frac{\partial^2}{\partial p^2} \ell(\beta|\mathbf{Y})\right] = -E\left[\frac{n}{\beta} - Z\right] = -E\left[-\frac{n}{\beta^2}\right] = \frac{n}{\beta^2}.$$

Therefore, the Jeffrey's prior distribution becomes

$$\pi(\beta) \propto \sqrt{|\mathcal{I}(\beta)|} \propto \frac{1}{\beta}.$$

The posterior distribution is then

$$\pi(\beta|\mathbf{Y}) \propto \beta^n e^{-\beta \sum_{k=1}^{n} Y_k} \cdot \frac{1}{\beta} = \beta^{n-1} e^{-\beta \sum_{k=1}^{n} Y_k},$$

which implies the posterior distribution is $\beta|\mathbf{Y} \sim \text{Gamma}(n, n\overline{Y})$. Since we recognize the posterior distribtion as a Gamma, we can easily report posterior point estimates as the mean, median, mode, etc. of the posterior distribution. Also, quantiles and HPD intervals can be computed in the normal fashion.

Problem 2:

(a) The distribution of $Y = X_1 + X_2$ is $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

(b) Assume $X_1 \sim N(3, 0.8)$ and $X_2 \sim N(-2.7, 0.23)$. We sample $10,000$ values each for X_1 and X_2 and compute $Y = X_1 + X_2$, i.e. we obtain $10,000$ samples of Y. The histogram below shows the probability of an outcome of Y based on the sample. We have also over-plotted a black kernel density estimate based on the outcomes of Y, as well as a red hyphened line

which is the true density of Y.

Histogram of Y
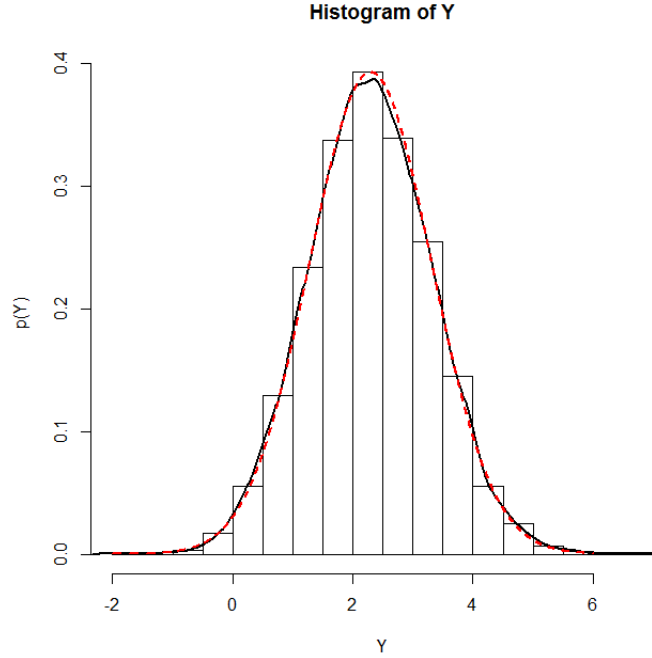
Here we can see that the kernel density estimate, i.e. plotting $Y = X_1 + X_2$, is almost indistinguishable from the true density of Y, i.e. $N(2.3, 1.03)$. Furthermore, the mean of Y was 2.303 and the variance of Y was 1.032. Both of these point estimates are very close to the true mean of 2.3 and true variance 1.03.

(c) To obtain the distribution of $Y = X_1 + X_2$, we can perform convolution. For $y \in \mathbb{R}$,

$$P(Y \leq y) = P(X_1 + X_2 \leq y) = \int_{-\infty}^{\infty} P(X_1 + x \leq y | X_1 = x) f_{X_2}(x) dx$$

$$= \int_{-\infty}^{\infty} P(X_1 \leq y - x) f_{X_2}(x) dx = \int_{-\infty}^{\infty} F_{X_1}(y - x) f_{X_2}(x) dx$$

where F_{X_1} denotes the CDF of X_1 and f_{X_2} denotes the PDF of X_2. Differentiating both sides will result in the PDF of Y, which is

$$f_Y(y) = \frac{\partial}{\partial y} \int_{-\infty}^{\infty} F_{X_1}(y - x) f_{X_2}(x) dx = \int_{-\infty}^{\infty} f_{X_1}(y - x) f_{X_2}(x) dx.$$

However, this integral is intractable in this case, and so we turn to Monte Carlo techniques. Notice that this integral can be thought of as an expected value, i.e.

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x) f_{X_2}(x) dx = E_{X_2}\big[f_{X_1}(y - x) \big]$$

and so by the law of large numbers,

$$f_Y(y) \approx \frac{1}{N} \sum_{k=1}^{N} f_{X_1}(y - x_k)$$

4

where $x_1, ..., x_k$ are samples from f_{X_2}. By Monte Carlo simulations, we obtain the graph in part (b) over-plotted with a blue line:

Histogram of Y
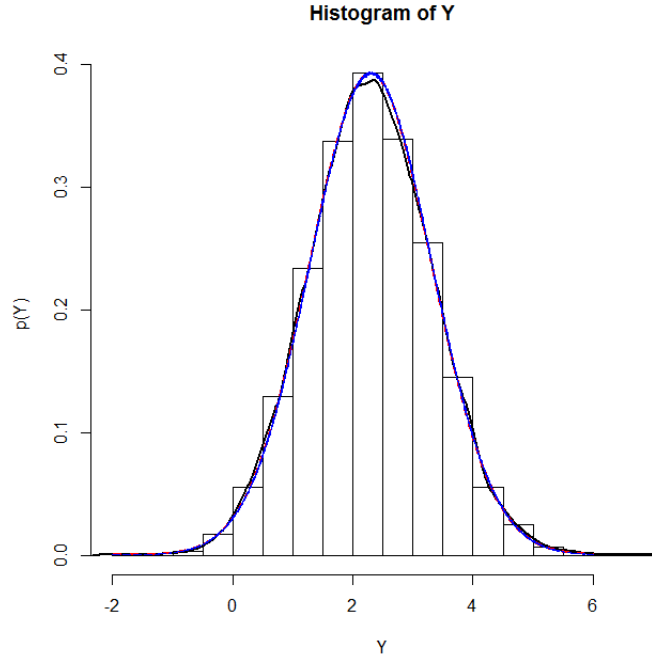
Here we can see that the convolution technique seemed to approximate the true density very well, in fact the two lines are almost indistinguishable in this situation.

(d) Now that we see these techniques appear to estimate the sum of random variables pretty well, consider $X_1 \sim \text{Gamma}(\alpha_1, \beta_1)$ and $X_2 \sim \text{Gamma}(\alpha_2, \beta_2)$. We are interested in the distribution of $Y = X_1 + X_2$. Since β_1 can differ from β_2, the distribution f_Y can be difficult to obtain analytically. So, we turn to Monte Carlo techniques. To illustrate, assume $X_1 \sim \text{Gamma}(2.3, 3)$ and $X_2 \sim \text{Gamma}(2.8, 2.2)$. We perform the same techniques as in parts (b)

and (c) of this problem. The results can be seen below:

Histogram of Y
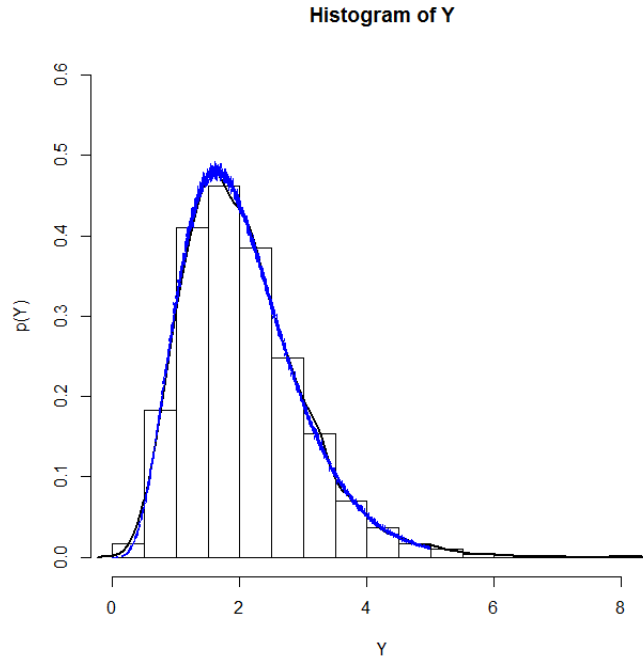
Again, the technique in part (b) is the black line, which is hidden under the blue line that represents the technique in part (c). For both, though, we see that the estimations appear to approximate the realizations of Y very well.

Problem 3:

(a) The game of craps is played in the following way. A player will roll two dice. If the sum of the two dice is a $2, 3$, or 12, the player loses the game. However, if the sum of the dice is a 7 or 11, then the player wins. If the sum of the dice is none of these values, then the value rolled is the players 'point'. Now the player repeatedly rolls the dice until he rolls his point or a 7. If the point is rolled again before a 7, then the player wins; otherwise he loses.

(b) Your initial bet will be 10 dollars. The return on any bet will be the amount that you bet, i.e. if you bet 10 dollars and win, then you win 10 dollars. You are going to play exactly 10 games of craps. If you lose on the previous game, your next bet requires you to "double down", that is, if on one game you bet X dollars and lose, on the next game you have to be $2X$ dollars. If you win on the previous game, your next bet will be 10 dollars.

(c) Using Monte Carlo simulation, we simulated $100,000$ outcomes of earnings after playing 10 games of craps with the strategy in part (b). The results came back of about negative $5 of expected earnings. Therefore, we conclude that this does not seem to be a profitable betting

strategy. The empirical distribution of earnings based on the $100,000$ data points is

Histogram of earnings
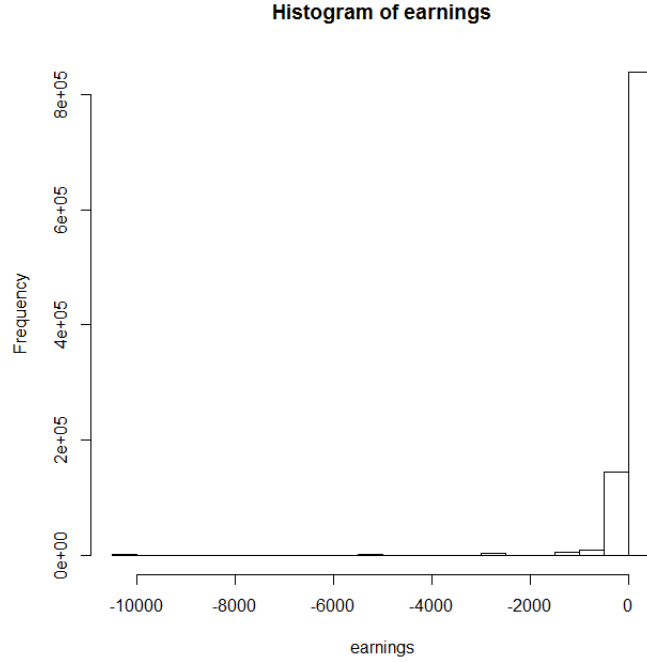
Althought it appears that most of the time we are profitable, notice that sometimes we lose a lot of money. The maximum earnings is $100, but we see that we can also lose up to $10,230$. Not a good idea in my opinion!

Problem 4:

(a) The parameter space for $\mathbf{p} = (p_1, p_2, p_3, p_4)$, the probability vector of the die that was rolled is

$$\Theta = \left\{ \begin{pmatrix} 1/3 \\ 1/3 \\ 1/6 \\ 1/6 \end{pmatrix}, \begin{pmatrix} 1/6 \\ 1/3 \\ 1/3 \\ 1/6 \end{pmatrix}, \begin{pmatrix} 1/6 \\ 1/6 \\ 1/6 \\ 1/2 \end{pmatrix} \right\} = \{\mathbf{p}_A, \mathbf{p}_B, \mathbf{p}_C\}.$$

(b) We first must calculate values of $L(\cdot|\mathbf{X})$ for the different possible dice rolled, where $\mathbf{X}$ is the outcomes observed. Notice,

$$L(A|\mathbf{X}) = \binom{29}{5} \left(\frac{1}{3}\right)^5 \binom{24}{11} \left(\frac{1}{3}\right)^{11} \binom{13}{6} \left(\frac{1}{6}\right)^6 \binom{7}{7} \left(\frac{1}{6}\right)^7 = 9.05 \cdot 10^{-4}$$

$$L(B|\mathbf{X}) = \binom{29}{5} \left(\frac{1}{6}\right)^5 \binom{24}{11} \left(\frac{1}{3}\right)^{11} \binom{13}{6} \left(\frac{1}{3}\right)^6 \binom{7}{7} \left(\frac{1}{6}\right)^7 = 1.81 \cdot 10^{-3}$$

$$L(C|\mathbf{X}) = \binom{29}{5} \left(\frac{1}{6}\right)^5 \binom{24}{11} \left(\frac{1}{6}\right)^{11} \binom{13}{6} \left(\frac{1}{6}\right)^6 \binom{7}{7} \left(\frac{1}{2}\right)^7 = 3.02 \cdot 10^{-5}.$$

Taking the maximum, we conclude that $\widehat{\mathbf{p}}_{MLE} = \mathbf{p}_B = (1/6, 1/3, 1/3, 1/6)$.

7

(c) To create a size α test of $H_0\colon \mathbf{p} = \mathbf{p}_C$ versus $H_1\colon \mathbf{p} \neq \mathbf{p}_C$, we considered a likelihood ratio test, i.e. our test statistic is

$$\Lambda = \frac{L(C|\mathbf{X})}{L(\widehat{\mathbf{p}}_{MLE}|\mathbf{X})},$$

with the decision rule of $I(\Lambda < c_\alpha)$, i.e. reject H_0 if $\Lambda < c_\alpha$ and fail to reject H_0 if $\Lambda \geq c_\alpha$, where c_α is chosen such that $P(\Lambda < c_\alpha) = \alpha$. Rather than trying to compute the proper c_α, we turn to Monte Carlo. We simulate $10,000$ possible outcomes of 29 rolls under dice C and recompute the MLE to obtain a sample $\Lambda^{(s)}$, for $s = 1, ..., 10000$. Then, we calculate what proportion of these fall less than Λ, which was 0.0024. Comparing this p-value to α, we reject if $0.0024 < \alpha$. This appears that we reject H_0 under any reasonable choice of α.

APPENDIX

```
################################################
################################################
################################################
#####
#####                 Chase Joyner
#####                882 Homework 2
#####              September 28, 2016
#####
################################################
################################################
################################################


#############################
##### Problem 2 Code #####

## Part b ##
mu1 = 5
mu2 = -2.7
sigma1sq = 0.8
sigma2sq = 0.23
n1 = n2 = 10000
X1 = rnorm(n1, mu1, sqrt(sigma1sq))
X2 = rnorm(n2, mu2, sqrt(sigma2sq))
Y = X1 + X2
hist(Y, prob = TRUE, xlab = "Y", ylab = "p(Y)")
lines(density(Y), lwd = 2)
sigmasq = sigma1sq + sigma2sq
mu = mu1 + mu2
tmp = seq(-2, 6, by = 0.001)
ptmp = 1 / sqrt(2*pi*sigmasq) * exp(-1/(2*sigmasq)*(tmp - mu)^2)
lines(tmp, ptmp, col = 2, lty = 2, lwd = 2)
mean(Y)
var(Y)

## Part c ##
mu1 = 5
mu2 = -2.7
sigma1sq = 0.8
sigma2sq = 0.23
y = seq(-2, 6, by = 0.001)
N = length(y)
dY = rep(-99, N)
for(i in 1:N){
```

```r
        X2.samp = rnorm(N, mu2, sqrt(sigma2sq))
        dY[i] = mean(dnorm(y[i]-X2.samp, mu1, sqrt(sigma1sq)))
}
lines(y, dY, col = "blue", lty = 5)


## Part d ##
alpha1 = 2.3
beta1 = 3
alpha2 = 2.8
beta2 = 2.2
n1 = n2 = 10000
X1 = rgamma(n1, alpha1, beta1)
X2 = rgamma(n2, alpha2, beta2)
Y = X1 + X2
hist(Y, prob = TRUE, xlab = "Y", ylab = "p(Y)", ylim = c(0,0.6))
lines(density(Y), lwd = 2)
y = seq(0, 5, by = 0.001)
N = length(y)
dY = rep(-99, N)
for(i in 1:N){
        X2.samp = rgamma(N, alpha2, beta2)
        dY[i] = mean(dgamma(y[i]-X2.samp, alpha1, beta1))
}
lines(y, dY, col = "blue", lty = 2)


x1 = rgamma(10000, alpha1, beta1)
x2 = rgamma(10000, alpha2, beta2)

hist(x1, prob = TRUE)
windows()
hist(x2, probe = TRUE)


##############################
##### Problem 3 Code #####

## Part c ##
playcraps = function(){
        num = append(1:6, 5:1)
        first.roll = sample(2:12, 1, prob = num / 36)
        if(first.roll %in% c(2,3,12)){
                return(0)
        } else if(first.roll %in% c(7,11)){
                return(1)
        } else{
```

```r
                        point = first.roll
                        next.roll = sample(2:12, 1, prob = num / 36)
                        while(next.roll != point && next.roll != 7){
                                next.roll = sample(2:12, 1, prob = num / 36)
                        }
                        if(next.roll == point){
                                return(1)
                        }
                        return(0)
                }
        }


n = 10000
res = rep(-99, n)
for(i in 1:n){
        res[i] = playcraps()
}
mean(res)          # Should be around 49%.  Checks out.


sims = 1000000   # Number of times simulated
n = 10            # Number of games to play
i.bet = 10        # Initial bet of 10 dollars
earnings = rep(0, sims)

for(t in 1:sims){
        bet = i.bet
        tmp = 0
        for(i in 1:n){
                res = playcraps()
                if(res == 0){
                        tmp = tmp - bet
                        bet = 2*bet
                }
                if(res == 1){
                        tmp = tmp + bet
                        bet = i.bet
                }
        }
        print(t)
        earnings[t] = tmp
}
mean(earnings)
hist(earnings)
```

```
#############################
##### Problem 4 Code #####

## Part c ##
elk = function(x,p){
        val = choose(29,x[1])*p[1]^x[1]*choose(29-x[1],x[2])*p[2]^x[2]
                *choose(29-x[1]-x[2],x[3])*p[3]^x[3]*p[4]^x[4]
        return(val)
}

X = c(5,11,6,7)
LpA = elk(X,c(1/3,1/3,1/6,1/6))
LpB = elk(X,c(1/6,1/3,1/3,1/6))
LpC = elk(X,c(1/6,1/6,1/6,1/2))

test.stat = LpC / max(LpA, LpB, LpC)

sims = 10000
stats = rep(-99, sims)
pA = c(1/3,1/3,1/6,1/6)
pB = c(1/6,1/3,1/3,1/6)
pC = c(1/6,1/6,1/6,1/2)
for(i in 1:sims){
        rolls = sample(1:4, 29, replace = TRUE, prob = pC)
        ones = length(which(rolls == 1))
        twos = length(which(rolls == 2))
        threes = length(which(rolls == 3))
        fours = length(which(rolls == 4))
        X = c(ones,twos,threes,fours)
        mle = max(elk(X,pA),elk(X,pB),elk(X,pC))
        stats[i] = elk(X,pC) / mle
}
mean(stats <= test.stat)
```