Analysis of successful putts from various distances of professional golfers

We are analysis successful putts from various distances of professional golfers. The data is given to us in a form of distance, no of tries for that distance, and number of successes for that distance.

It was chosen to compare 4 generalised linear models where for both we have considered the response variable to be the number of successes, which naturally fits binomial distribution:

1. $y_i \sim Bin(n, p_i), logit(p_i) = \alpha$
2. $y_i \sim Bin(n, p_i), logit(p_i) = \alpha + \beta x_i$
3. $y_i \sim Bin(n, p_i), logit(p_i) = \alpha + \beta_1 x_i + \beta_2 x_i^2$
4. $y_i \sim Bin(n, p_i), logit(p_i) = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3$

Full Code can be found in Appendix where JAGS code for Models can be found at golfers_bin1.model, golfers_bin2.model. R code is at R Code.

Both models were created using three chains and samples were drawn using 10000 iterations.

For each of the models, we have generated DIC values and Table 1 contains the results.

Model	DIC
Model 1	2520
Model 2	365.9
Model 3	185.8
Model 4	155.4

Table 1. DIC Values

In this case the clear choice is Model 4 due to lowest DIC value.

Mean values for each of the parameters are:

$$\alpha = -0.91774648$$

$$\beta_1 = -0.12141928$$

$$\beta_2 = 0.01365918$$

$$\beta_3 = -0.00156764$$

The next step was to run cross-correlation. We can see on Figure 1 that there is a strong negative correlation between α and β_2, and between β_1 and β_3. One approach to fix this is to potentially do reparameterization which is not done for this analysis due to the low DIC value.

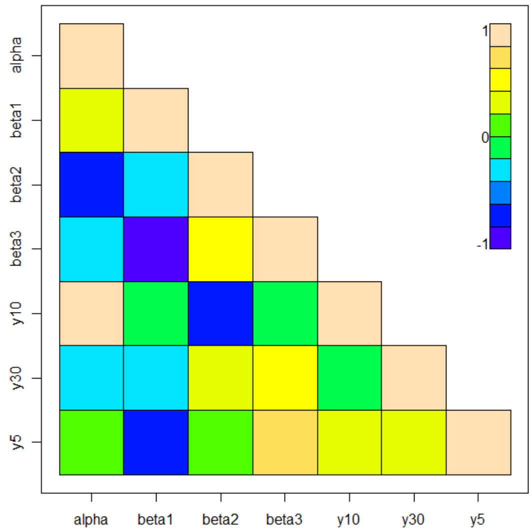

Figure 1. Cross Correlation Plot

Finally, we have completed estimates of proportion of success for 5, 10 and 30 feet distance attempts and the results are in the Table 2 including 95% credible interval.

Distance	Estimate	95% credible interval
5	0.65491896	(6.329279e-01, 0.676514972)
10	0.31418449	(2.931234e-01, 0.335689129)
30	0.00070438	(2.436157e-06, 0.004912875)

The estimates for 5 and 10 feet are slightly higher than the numbers we have from the input data (0.589 and 0.335) and are out of credible interval which brings us to believe that we could parameterize this model to improve it.

Appendix

R Code

```
require(rjags)

# the data
golfers_data=read.delim("golfers.dat")
golfers_data$ProportionSuccess = golfers_data$Successes / golfers_data$Tries


golfers.data = list(
  N=nrow(golfers_data),
  Distance=golfers_data$Distance,
  Tries=golfers_data$Tries,
  Successes = golfers_data$Successes)

# Models and sampling

golfers.model_bin1=jags.model("golfers_bin1.model", golfers.data, n.chains=3)
golfers.samps_bin1=coda.samples(golfers.model_bin1,variable.names=c("alpha", "y5", "y10",
"y30"),1e4)

golfers.model_bin2=jags.model("golfers_bin2.model", golfers.data, n.chains=3)
```

```
golfers.samps_bin2=coda.samples(golfers.model_bin2,variable.names=c("alpha","beta", "y5",
"y10", "y30"),1e4)

golfers.model_bin3=jags.model("golfers_bin3.model", golfers.data, n.chains=3)
golfers.samps_bin3=coda.samples(golfers.model_bin3,
                                variable.names=c("alpha","beta1", "beta2", "y5", "y10",
"y30"),1e4)

golfers.model_bin4=jags.model("golfers_bin4.model", golfers.data, n.chains=3)
golfers.samps_bin4=coda.samples(golfers.model_bin4,
                                variable.names=c("alpha","beta1", "beta2", "beta3",
                                                 "y5", "y10", "y30"),1e4)

# DIC
golfers.dic_bin1 = dic.samples(golfers.model_bin1, 1e4)
golfers.dic_bin2 = dic.samples(golfers.model_bin2, 1e4)
golfers.dic_bin3 = dic.samples(golfers.model_bin3, 1e4)
golfers.dic_bin4 = dic.samples(golfers.model_bin4, 1e4)

golfers.dic_bin1
golfers.dic_bin2
golfers.dic_bin3
golfers.dic_bin4

# Sumamry statistics

golfers.summary_bin1 <- summary(golfers.samps_bin1)
golfers.means_bin1 <- golfers.summary_bin1$statistics[,"Mean"]

golfers.summary_bin2 <- summary(golfers.samps_bin2)
golfers.means_bin2 <- golfers.summary_bin2$statistics[,"Mean"]

golfers.summary_bin3 <- summary(golfers.samps_bin3)
golfers.means_bin3 <- golfers.summary_bin3$statistics[,"Mean"]

golfers.summary_bin4 <- summary(golfers.samps_bin4)
golfers.means_bin4 <- golfers.summary_bin4$statistics[,"Mean"]

# Print means
golfers.means_bin1
golfers.means_bin2
golfers.means_bin3
golfers.means_bin4

# Print quantiles
golfers.summary_bin1$quantiles
golfers.summary_bin2$quantiles
golfers.summary_bin3$quantiles
golfers.summary_bin4$quantiles

par(mfrow=c(1,1))

crosscorr(golfers.samps_bin4)
crosscorr.plot(golfers.samps_bin4)
```

golfers_bin1.model

```
model{
    for (i in 1:N){
            Successes[i] ~ dbin(p[i], Tries[i])
            logit(p[i]) <- alpha
    }

    alpha ~ dnorm(0, 0.0001)
    p_temp5 <- alpha
```

```
    p_temp10 <- alpha
    p_temp30 <- alpha
    y5 <- 1 / (1 + exp(-p_temp5))
    y10 <- 1 / (1 + exp(-p_temp10))
    y30 <- 1 / (1 + exp(-p_temp30))
}
```

golfers_bin2.model

```
model{
    for (i in 1:N){
            Successes[i] ~ dbin(p[i], Tries[i])
            logit(p[i]) <- alpha + beta*(Distance[i]-mean(Distance[]))
    }

    alpha ~ dnorm(0, 0.0001)
    beta ~ dnorm(0, 0.0001)
    p_temp5 <- alpha + beta*(5-mean(Distance[]))
    p_temp10 <- alpha + beta*(10-mean(Distance[]))
    p_temp30 <- alpha + beta*(30-mean(Distance[]))
    y5 <- 1 / (1 + exp(-p_temp5))
    y10 <- 1 / (1 + exp(-p_temp10))
    y30 <- 1 / (1 + exp(-p_temp30))
}
```

golfers_bin3.model

```
model{
    for (i in 1:N){
            Successes[i] ~ dbin(p[i], Tries[i])
            logit(p[i]) <- alpha + beta1*(Distance[i]-mean(Distance[])) +
beta2*pow(Distance[i]-mean(Distance[]),2)
    }

    alpha ~ dnorm(0, 0.0001)
    beta1 ~ dnorm(0, 0.0001)
    beta2 ~ dnorm(0, 0.0001)
    p_temp5 <- alpha + beta1*(5-mean(Distance[])) + beta2*pow(5-mean(Distance[]),2)
    p_temp10 <- alpha + beta1*(10-mean(Distance[])) + beta2*pow(10-mean(Distance[]),2)
    p_temp30 <- alpha + beta1*(30-mean(Distance[])) + beta2*pow(30-mean(Distance[]),2)
    y5 <- 1 / (1 + exp(-p_temp5))
    y10 <- 1 / (1 + exp(-p_temp10))
    y30 <- 1 / (1 + exp(-p_temp30))
}
```

golfers_bin4.model

```
model{
    for (i in 1:N){
            Successes[i] ~ dbin(p[i], Tries[i])
            logit(p[i]) <- alpha + beta1*(Distance[i]-mean(Distance[])) +
beta2*pow(Distance[i]-mean(Distance[]),2) + beta3*pow(Distance[i]-mean(Distance[]),3)
    }

    alpha ~ dnorm(0, 0.0001)
    beta1 ~ dnorm(0, 0.0001)
    beta2 ~ dnorm(0, 0.0001)
    beta3 ~ dnorm(0, 0.0001)

    p_temp5 <- alpha + beta1*(5-mean(Distance[])) + beta2*pow(5-mean(Distance[]),2) +
beta3*pow(5-mean(Distance[]),3)
    p_temp10 <- alpha + beta1*(10-mean(Distance[])) + beta2*pow(10-mean(Distance[]),2) +
beta3*pow(10-mean(Distance[]),3)
    p_temp30 <- alpha + beta1*(30-mean(Distance[])) + beta2*pow(30-mean(Distance[]),2) +
beta3*pow(30-mean(Distance[]),3)
    y5 <- 1 / (1 + exp(-p_temp5))
    y10 <- 1 / (1 + exp(-p_temp10))
    y30 <- 1 / (1 + exp(-p_temp30))
}
```