Stat 532 Assignment 2

Kenny Flagg

September 11, 2015

1.
   - Gelman et al define probabilities as numbers associated with outcomes which are "nonnegative, additive over mutually exclusive outcomes, and sum to 1 over all possible mutually exclusive outcomes" (BDA3, §1.5, p. 11).
   - People tend to use probability to describe their level of certainty about an outcome. Common uses that most people are exposed to include the probability of winning a sweepstakes or the chance that it will rain tomorrow. People don't generally think of a 1 in 250 chance of winning a drawing as "if I entered a drawing like this 250 times I should expect to win once," they interpret is as "I am almost certainly not going to win."

2. Jordan's criticism (which he found in the Wikipedia article about the likelihood principle) was that the likelihood principle ignores the context of the data. It implies that identical data should lead to the same conclusions, but this is not always appropriate.

   Suppose we are have a coin that we suspect is weighted to come up heads more often than tails. We a null hypothesis that $\theta = Pr(\text{heads}) <= 0.5$ and we seek evidence that $\theta > 0.5$. We could devise two different experiments to test these hypotheses:

   (a) Flip the coin 6 times and record y, the number of heads that appear. The model is $y \sim Binom(6, \theta)$ with mass function

   $$p(y|\theta) = \binom{6}{y}\theta^y(1-\theta)^{6-y}; y = 0, 1, 2, 3, 4, 5, 6; 0 < \theta < 1.$$

   (b) Flip the coin until a tails appears and record x, the number of heads that appear before the first tails. This model is $x \sim NBinom(1, \theta)$ and the probability mass function is

   $$p(x|\theta) = \binom{x}{0}\theta^x(1-\theta); x = 0, 1, 2, 3, \dots; 0 < \theta < 1.$$

   In each experiment, we could observe 5 heads and 1 tails. In this case, the likelihood function for (a) would be

   $$L(\theta|y = 5) = 6\theta^5(1-\theta)$$

   and the likelihood function for (b) would be

   $$L(\theta|x = 5) = \theta^5(1-\theta).$$

   Since these are proportional to each other, they both result in the same maximum likelihood point estimate, $\hat{\theta} = \dfrac{5}{6}$.

   If we compute p-values from the randomization distribution, we get

   $$Pr(y \geq 5|p = 0.5) = 0.1094$$

and

$$Pr(x \geq 5|p = 0.5) = 0.03125.$$

The first result is, at best, very weak evidence that the coin is biased toward heads. The second is strong evidence of a bias towards heads. The difference comes from the design of the experiments – In the first experiment, the tails could occur anywhere in the sequence of coin flips. In the second, the tails must appear after a run of 5 heads, a much less probable event. Strict adherence to the likelihood principle ignores the different ways that data can arise.

One could counter this argument by pointing out that the sampling distribution of $\hat{p}$ differs between the two experiments. The estimate from (a) is more variable than the estimate from (b). Additionally, the likelihoods are not identical because they differ by a factor of 6. However, this is problematic when Bayesian methods are used because this factor will disappear when inverting to find the posterior distribution. A Bayesian analysis of these experiments using the same prior distributions would result in identical posterior distributions and thus incorrectly lead to the same conclusions.

3. A likelihood function tells us what parameter values are reasonable, given the single dataset at hand. Since the data vary from sample to sample, any method of finding a point estimate is expected to give different results from different samples. The estimates will deviate from the true parameter value, occasionally having a very large deviation. (I think my plots for problem 10 illustrate this well.) However, when the sample size is large the maximum likelihood estimator has a distribution that is approximately normal with a mean equal to the value of the parameter. This property of having a friendly sampling distribution makes maximizing the likelihood a reasonble thing to do when faced with sampling variability.

4. A likelihood function is a relative measurement of how well a stated value fits with the specific set of data that were observed.

   For example, suppose we have an urn that we know contains 12 marbles, and we want to estimate how many of the marbles are gold without pouring out all of the marbles. We could randomly draw a few of the marbles and count how many of these that are gold. We'll use x to denote the number of gold marbles the we drew, and we'll call the total number of gold marbles θ.

   Let's say we choose 5 marbles, and 1 is gold. If we knew that there were a total of 2 gold marbles, we could find the probability $Pr(x|\theta = 2)$ of choosing any number of gold marbles. It turns out that $Pr(x = 1|\theta = 2) = 0.53$. If we knew that the total number of gold marbles was 3, we could do the same thing and find that $Pr(x = 1|\theta = 3) = 0.48$. This means that a the claim that there are 2 gold marbles fits with our observation better than the claim that there are 3 gold marbles.

   The numbers 0.53 and 0.48 are values of the likelihood function. We represent it as $L(\theta|x)$ to emphasize that is comparing possible values of θ; the value of x is fixed to be what was observed. It seems reasonable to use the θ with the largest likelihood value as our estimate, but we may not always be correct. To get a better sense of what our estimate should be, we can consider this estimate probabilistically. Values of the likelihood are not probabilities (notice that $0.53 + 0.48 > 1$, and any number θ bigger than the observed x could be possible), but we can theoretically find the distribution of the values this estimate could take for our given x.

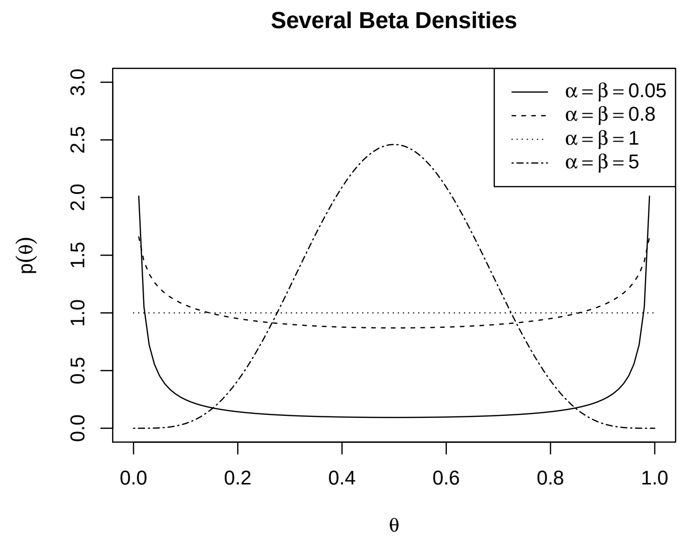
5. Wilson's example was that, in a study meant to estimate the survival rate of ducklings, 24 ducklings were observed and all 24 survived until the conclusion of the study. As a result, the estimated survival rate is $\hat{p} = 1$ with a standard error of 0. This represents a failure the frequentist paradigm that can easily be remedied in a Bayesian setting.

   The fully data-driven analysis does not reflect all available information. The frequentist analysis leads to the conclusion that the proportion of ducklings who survive is exactly 1. However, any reasonable person knows that ducks are not immortal and indestructible; their survival rate in any length of time is less than 1. Even if the study period was too short for the expected number of deaths in a group of 24 ducklings to exceed 1, the researchers had a reason to be investigating the survival rate. This indicates the existence of a priori knowledge that not all of the ducklings should survive.

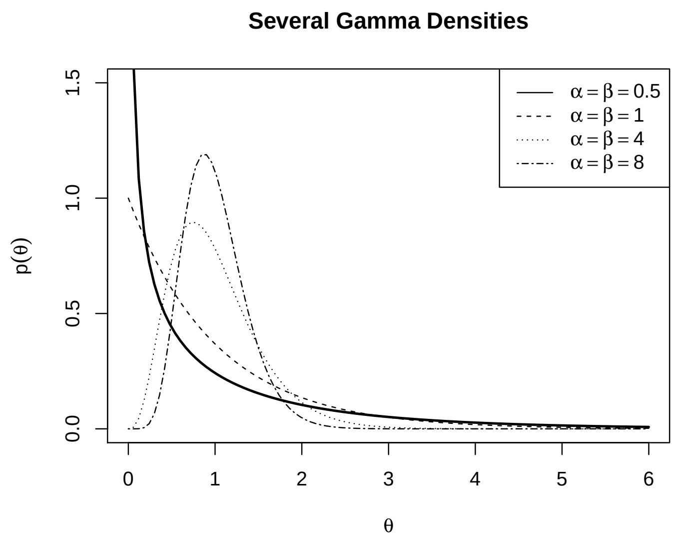
Such an error would not occur so easily if Bayesian methods were used. The Bayesian model would incorporate the prior knowledge that the survival rate must be strictly less than 1. In the absence of any specific information about what the rate might be, the analysis could still be performed. The result would be more precise, updated knowledge about the rate of survival for the ducks under the study conditions.

Frequentists would counter that their methods are accurate in the long run and bound to lead to false conclusions a small proportion of the time. It is unnecessarily limiting to require that this one study must only be considered in the context of many repetitions of identical studies. Bayesian methods provide a formal way to combine existing knowledge with the data at hand and discuss certainty, not error rates. The individual study is allowed to have merit.

6. Below are several beta density curves with $\alpha = \beta$. When $\alpha \neq \beta$, increasing α increases left-skew and increasing β increases right-skew.



7. I had not realized that R's default parameterization for the Gamma distribution uses a rate parameter instead of a scale parameter. Fortunately this agrees with the parameterization in BDA3 (where the rate parameter is called an inverse scale parameter). It makes me wonder how many times I've incorrectly used a parameterization that I did not intend to use. I've plotted several gamma density curves below.



The inverse scale parameterization of the Gamma distribution forms a conjugate prior when used to model Poisson rates and Normal precisions (inverse variances), so it is often a default choice in these situations.

8. (a) The posterior probability of infection is

$$\begin{aligned} Pr(\text{infection}|\text{test }+) &= \frac{Pr(\text{test }+|\text{infection})Pr(\text{infection})}{Pr(\text{test }+|\text{infection})Pr(\text{infection}) + Pr(\text{test }+|\text{no infection})Pr(\text{no infection})} \\ &= \frac{0.92Pr(\text{infection})}{0.92Pr(\text{infection}) + 0.08Pr(\text{no infection})} \end{aligned}$$

If the first doctor is correct, this works out to

$$Pr(\text{infection}|\text{test }+) = \frac{(0.92)(0.05)}{(0.92)(0.05) + (0.08)(0.95)} = 0.377$$

so there is a 37.7% chance that the patient is infected.
If the second doctor is right,

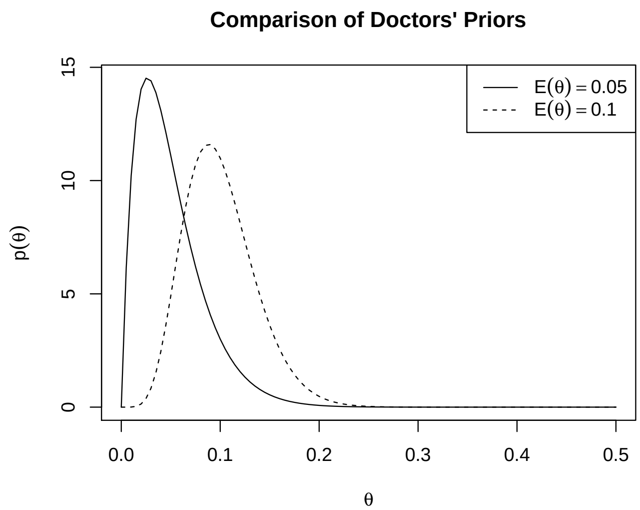
$$Pr(\text{infection}|\text{test }+) = \frac{(0.92)(0.1)}{(0.92)(0.1) + (0.08)(0.9)} = 0.561$$

and the patient has a 56.1% probability of being infected.

The second doctor believed the infection rate to be twice as large as what the first doctor believed. As a result, the second doctor would consider the patient as slightly more likely to be infected than not, while the first doctor would conclude that the patient is most likely not infected.

I see this as illustrating the Bayesian interpretation of probability as describing the state a knowledge about something unobserved. If we could observe the infection status without measurement error (a perfect test) then we would definitively know if the patient is infected or not. Since the patient's status is unknown, the doctors must incorporate their previous knowledge about the infection and they come to different conclusions despite both observing the same test result. If we think that objectivity means that doctors should follow the likelihood principle and reach the same conclusions given the same data, then this example could support an argument for the use of uninformative priors when experts have disagreement.

(b) Let $y = 1$ if the patient tests positive, 0 otherwise. I will use $Beta(2, 38)$ for the first doctor's prior and $Beta(7, 63)$ for the second doctor's prior. These have similar variances (0.00116 and 0.00127) and their means agree with the doctors' opinions. The densities are plotted below.



If I knew something about how the doctors formed their opinions I might choose priors to reflect that, but since I do not have that information I treat the opinions as point estimates and assume that they are approximately equally precise.

The probability model for y is

$$Pr(y = 1|\theta) = 0.92\theta + 0.08(1 - \theta)$$

9. Let y be the number or girls in $n = 241945 + 251527 = 493472$ births. The model is $y|\theta \sim Binomial(493472, \theta)$. Laplace used the prior $\theta \sim Beta(1, 1)$. Then

$$p(y, \theta) = \left(\binom{493472}{y} \theta^y (1-\theta)^{493472-y} \right) \left(I_{(0,1)}(\theta) \right) = \binom{493472}{y} \theta^y (1-\theta)^{493472-y}$$

so

$$p(\theta|y) \propto \theta^y (1-\theta)^{493472-y}$$

and we see that the posterior is $\theta|y \sim Beta(y + 1, 493473 - y)$.

After observing $y = 241945$ female births, we make posterior inferences from the $Beta(241946, 251528)$ distribution. R tells me that

$$Pr(\theta \geq 0.5|y = 241945) = 1.14606 \times 10^{-42}.$$

Thus, rounding to any reasonable number of significant digits, we get that essentially

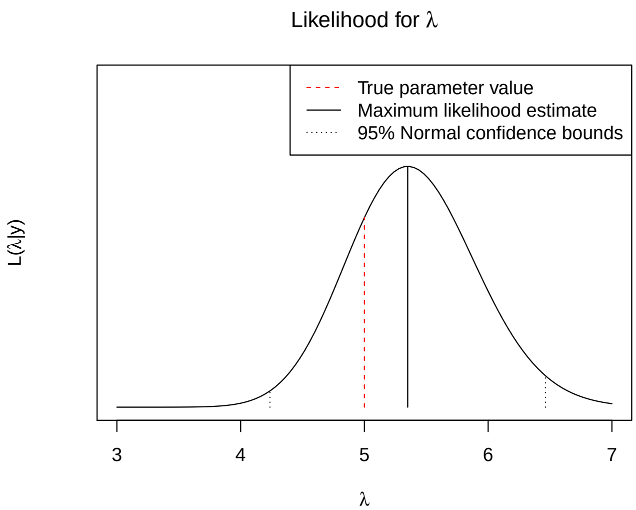
$$Pr(\theta < 0.5|y = 241945) = 1.$$

I wouldn't use the phrase "morally certain," but the evidence is overwhelming that fewer than half of the births around that time were females.

10. (a) 20 observations from $Poisson(\lambda = 5)$:

Mean	Var.	Min.	1st Qu.	Median	3rd Qu.	Max.
5.35	6.45	1.00	4.00	5.00	6.50	11.00

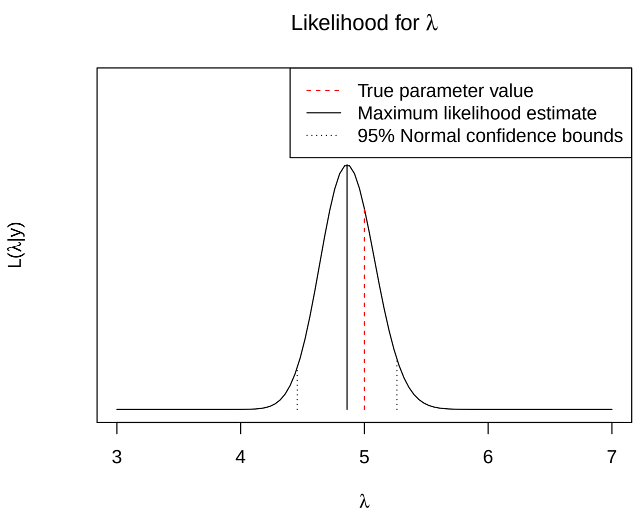
Maximum Likelihood Estimate: $\hat{\lambda} = \bar{y} = 5.35$, SE: $\frac{s}{\sqrt{n}} = 0.568$



(b) 100 observations from $Poisson(\lambda = 5)$:

Mean	Var.	Min.	1st Qu.	Median	3rd Qu.	Max.
4.86	4.22	0.00	3.00	5.00	6.00	11.00

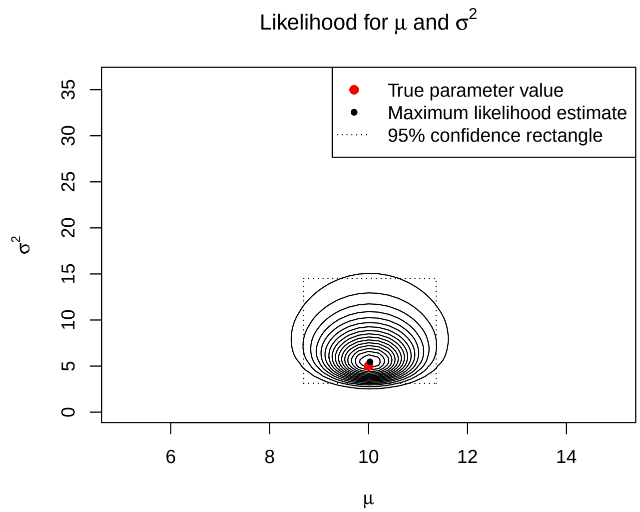
Maximum Likelihood Estimate: $\hat{\lambda} = \bar{y} = 4.86$, SE: $\frac{s}{\sqrt{n}} = 0.205$



(c) 15 observations from $N(\mu = 10, \sigma^2 = 5)$:

Mean	Var.	Min.	1st Qu.	Median	3rd Qu.	Max.
10.03	5.84	5.51	8.27	9.32	11.72	13.74

MLEs: $\hat{\mu} = \bar{y} = 10.03$, $SE(\bar{y}) = \dfrac{s}{\sqrt{n}} = 0.624$, $\hat{\sigma}^2 = \dfrac{\sum(y_i - \bar{y})^2}{n} = 5.452$

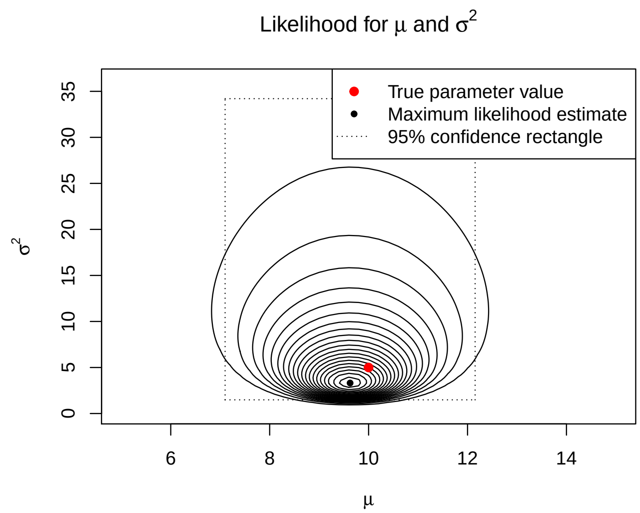


The confidence bounds are based on $\dfrac{\bar{y} - \mu}{s/\sqrt{n}} \sim t_{14}$ and $\dfrac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{14}$.

(d) 5 observations from $N(\mu = 10, \sigma^2 = 5)$:

Mean	Var.	Min.	1st Qu.	Median	3rd Qu.	Max.
9.63	4.14	7.30	8.88	9.26	9.83	12.86

MLEs: $\hat{\mu} = \bar{y} = 9.63$, $SE(\bar{y}) = \dfrac{s}{\sqrt{n}} = 0.91$, $\hat{\sigma}^2 = \dfrac{\sum(y_i - \bar{y})^2}{n} = 3.314$

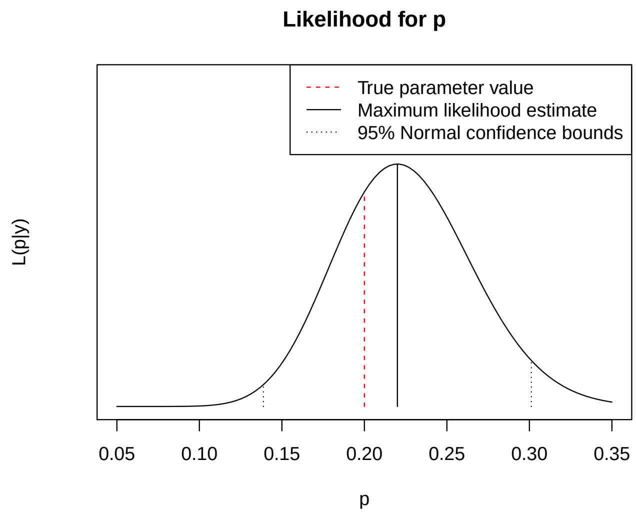


The confidence bounds are based on $\dfrac{\bar{y} - \mu}{s/\sqrt{n}} \sim t_4$ and $\dfrac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_4$.

(e) 1 observation from $Binomial(m = 100, p = 0.2)$:

Observed $y = 22$ successes in $m = 100$ trials

Maximum Likelihood Estimate: $\hat{p} = \dfrac{y}{m} = 0.22$, SE: $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 0.04142$

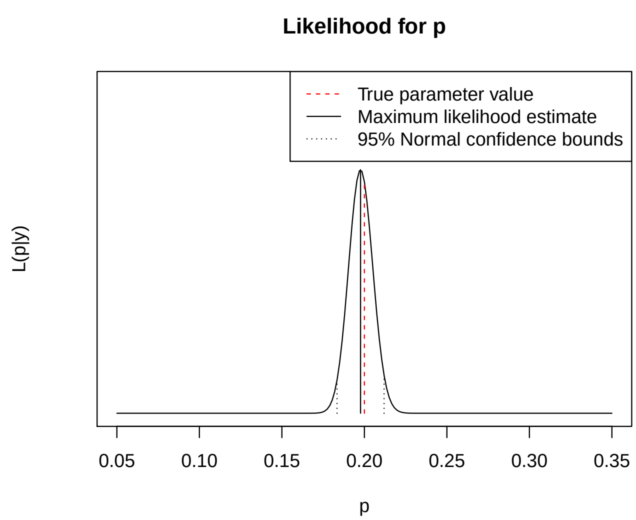


(f) 30 observations from $Binomial(m = 100, p = 0.2)$:

Mean	Var.	Min.	1st Qu.	Median	3rd Qu.	Max.
19.77	20.87	12.00	16.00	19.50	22.75	29.00

Observed $\sum y_i = 593$ successes in $nm = 3000$ trials

Maximum Likelihood Estimate: $\hat{p} = \dfrac{\sum y_i}{nm} = 0.198$, SE: $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{nm}} = 0.00727$



(g) The plots show very nicely how larger sample sizes lead to more precise inferences, as the likelihoods are much narrower for the larger samples. I plotted the pairs on the same scale to emphasize this.

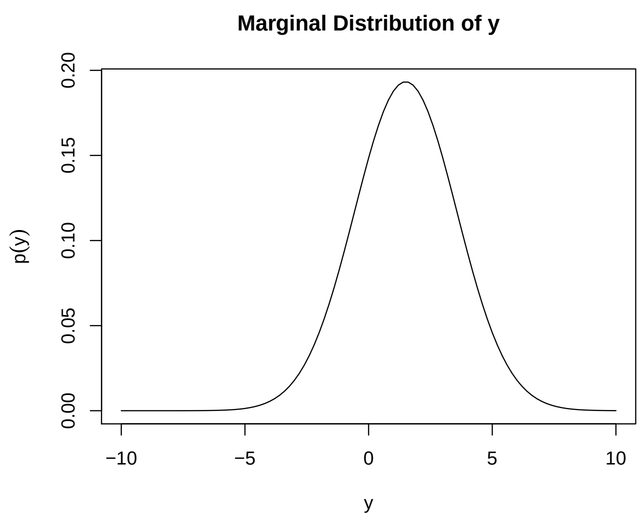
It is also interesting to observe that the true parameter can have a noticeably lower likelihood than the MLE due to sampling variability (especially noticeable in part (a)). Another thing to note is that the likelihood is not constant on the boundaries of the usual frequentist confidence sets. These are constructed to be accurate through repeated sampling, not to have endpoints that are equally "likely" given one data set.

11. (a) We have $y \sim N(\theta, \sigma^2)$ with $Pr(\theta = 1) = Pr(\theta = 2) = \frac{1}{2}$. If $\sigma = 2$,

$$p(y, \theta) = \left(\frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(y - \theta^2)}{8}\right)\right)\left(\frac{1}{2}\right) = \frac{1}{4\sqrt{2\pi}} \exp\left(-\frac{(y-\theta)^2}{8}\right); \theta = 1, 2$$

so the marginal density of y is

$$p(y) = \sum_{\theta=1}^{2} p(y, \theta) = \frac{1}{4\sqrt{2\pi}} \exp\left(-\frac{(y-1)^2}{8}\right) + \frac{1}{4\sqrt{2\pi}} \exp\left(-\frac{(y-2)^2}{8}\right)$$
$$= \frac{1}{4\sqrt{2\pi}}\left(e^{-\frac{(y-1)^2}{8}} + e^{\frac{(y-2)^2}{8}}\right).$$



(b) The posterior probability is

$$Pr(\theta = 1|y = 1) = \frac{p(y = 1, \theta = 1)}{p(y = 1)} = \frac{\frac{1}{4\sqrt{2\pi}} e^{-\frac{(1-1)^2}{8}}}{\frac{1}{4\sqrt{2\pi}}\left(e^{-\frac{(1-1)^2}{8}} + e^{-\frac{(1-2)^2}{8}}\right)}$$
$$= \frac{1}{1 + e^{-\frac{1}{8}}} \approx 0.5312.$$

(c) For any $\sigma > 0$, the posterior distribution is

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} = \frac{\frac{1}{2\sigma\sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}}{\frac{1}{2\sigma\sqrt{2\pi}}\left(e^{-\frac{(y-1)^2}{2\sigma^2}} + e^{-\frac{(y-2)^2}{2\sigma^2}}\right)}$$
$$= \frac{e^{-\frac{(y-\theta)^2}{2\sigma^2}}}{e^{-\frac{(y-1)^2}{2\sigma^2}} + e^{-\frac{(y-2)^2}{2\sigma^2}}}; \theta = 1, 2$$

so as σ increases, each exponent approaches 0 and thus $p(\theta|y) \to \frac{1}{2}$ for $\theta = 1, 2$.

θ appears only in the exponent of the numerator, so as σ decreases, $-\dfrac{(y-\theta)^2}{2\sigma^2}$ decreases and the distribution is pulled towards y.

12.

$$\begin{aligned}
Pr(\text{identical}|2\text{ males}) &= \frac{Pr(2\text{ males}|\text{identical})Pr(\text{identical})}{Pr(2\text{ males}|\text{identical})Pr(\text{identical}) + Pr(2\text{ males}|\text{fraternal})Pr(\text{fraternal})} \\
&= \frac{Pr(\text{male})Pr(\text{identical})}{Pr(\text{male})Pr(\text{identical}) + Pr(\text{male})Pr(\text{male})Pr(\text{fraternal})} \\
&= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{300}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{300}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{125}\right)} \\
&= \frac{\frac{1}{300}}{\frac{1}{300} + \frac{1}{250}} = \frac{5}{11}
\end{aligned}$$

13. (a) If we consider the outcome of the roll to be an observable quantity that A has observed and B has not, then I_A contains the outcome of roll but I_B contains only the information that the die is fair. Then the probabilities assigned by A and B are

$$\begin{aligned}
P_A(6) &= \begin{cases} 1 & I_A = \text{“The roll is a 6.”} \\ 0 & I_A = \text{“The roll is not a 6.”} \end{cases} \\
P_B(6) &= \frac{1}{6}.
\end{aligned}$$

Since A has perfect knowledge of the outcome, A no longer considers the event to be random. However, lack of knowledge leads B to assign probabilities based on the physical process of the die roll.

(b) An individual with little knowledge of soccer, such as A or myself, might look up the number of FIFA member countries on Wikipedia and assume all of the 209 countries' teams are equally likely to qualify for and win the World Cup. Then I_A = “There are 209 countries that participate in FIFA” and A would assign the probability $P_A(\text{Brazil wins the World Cup}) = \dfrac{1}{209}$.

An avid follower like B would adjust the probability based on observed information about the players and the outcomes of matches in the qualifying round, as well as on the structure of the qualifying tournament. Perhaps, based on the historical record, B would assume that Brazil is guaranteed to be one of the 32 teams that qualify for the World Cup and has a better chance of winning than some other qualifying teams, and then assign a probability of $P_B(\text{Brazil wins the World Cup}) = p > \dfrac{1}{32}$. The exact value could vary a lot based on what information B observed from watching previous matches.

14. (b) If $y \sim Binomial(n, \theta)$ and $\theta \sim Beta(\alpha, \beta)$, then $\theta|y \sim Beta(\alpha + y, \beta + n)$ so the posterior mean is $E(\theta|y) = \dfrac{\alpha + y}{\alpha + \beta + n}$.

If $\dfrac{\alpha}{\alpha+\beta} < \dfrac{y}{n}$ then

$$\begin{aligned}
& n\alpha < y(\alpha+\beta) \\
\implies & n\alpha + ny < y(\alpha+\beta) + ny \\
\implies & n(\alpha+y) < y(\alpha+\beta+n) \\
\implies & \frac{\alpha+y}{\alpha+\beta+n} < \frac{y}{n}.
\end{aligned}$$

Also,

$$\begin{aligned}
& n\alpha < y(\alpha+\beta) \\
\implies & \alpha(\alpha+\beta) + n\alpha < \alpha(\alpha+\beta) + y(\alpha+\beta) \\
\implies & \alpha(\alpha+\beta+n) < (\alpha+\beta)(\alpha+y) \\
\implies & \frac{\alpha}{\alpha+\beta} < \frac{\alpha+y}{\alpha+\beta+n},
\end{aligned}$$

so

$$\frac{\alpha}{\alpha+\beta} < \frac{\alpha+y}{\alpha+\beta+n} < \frac{y}{n}.$$

If $\frac{\alpha}{\alpha+\beta} > \frac{y}{n}$, the same steps (with the directions of the inequalities reversed) will establish

$$\frac{\alpha}{\alpha+\beta} > \frac{\alpha+y}{\alpha+\beta+n} > \frac{y}{n}.$$

Therefore, if $\frac{\alpha}{\alpha+\beta} \neq \frac{y}{n}$ then the posterior mean of θ is between these values.

(c) If $\theta \sim Unif(0,1) = Beta(1,1)$, then $\theta|y \sim Beta(y+1, n-y+1)$. The prior variance is

$$Var(\theta) = \frac{1}{12}$$

and the posterior variance is

$$Var(\theta|y) = \frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}.$$

Since the quadratic $(t+1)(n-t+1)$ is maximized by $t = \frac{n}{2}$,

$$Var(\theta|y) = \frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} \leq \frac{(\frac{n}{2}+1)^2}{(n+2)^2(n+3)}.$$

The rightmost expression is decreasing in n for $n \geq 0$. If any data is observed then $n \geq 1$, so therefore

$$Var(\theta|y) \leq \frac{(\frac{1}{2}+1)^2}{(1+2)^2(1+3)} = \frac{1}{16} < \frac{1}{12} = Var(\theta).$$

(d) The variance could increase if we use an informative prior and then observe contradictory data. If our prior is based on $\alpha = 9$ successes and $\beta = 1$ failures, and then we observe $y = 1$ success in another $n = 10$ trials, the prior variance is

$$Var(\theta) = \frac{(9)(1)}{(9+1)^2(9+1+1)} \approx 0.0082$$

but the posterior variance is

$$Var(\theta|y) = \frac{(9+1)(1+10-1)}{(9+1+10)^2(9+1+10+1)} \approx 0.0119.$$

15. (a) We have $\theta \in (0, \infty)$ and use the "non-informative" improper prior $p_\theta(\theta) = I_{(0,\infty)}(\theta)$. If we want to consider the transformation $\phi = \log(\theta)$ then the prior distribution of ϕ is

$$p_\phi(\phi) = p_\theta(e^\phi) \left| \frac{d\theta}{d\phi} \right| = I_{(0,\infty)}(e^\phi) e^\phi = I_{(\infty,\infty)}(\phi) e^\phi$$

which looks like we are expecting very large values of ϕ.

(b) The likelihood is

$$p(y|\theta) = \frac{e^{-\theta}\theta^y}{y!}$$

so

$$\begin{aligned}
\log(p(y|\theta)) &= -\theta + y\log(\theta) - \log(y!), \\
\frac{d}{d\theta}\log(p(y|\theta)) &= -1 + \frac{y}{\theta}, \\
\frac{d^2}{d\theta^2}\log(p(y|\theta)) &= -\frac{y}{\theta^2}.
\end{aligned}$$

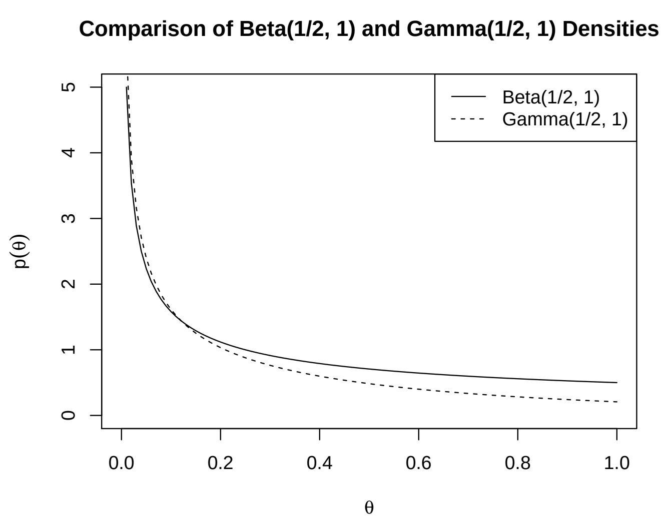
Then the Fisher information is

$$J(\theta) = -E\left(\frac{d^2}{d\theta^2}\log(p(y|\theta))|\theta\right) = -E\left(-\frac{y}{\theta^2}|\theta\right) = \frac{1}{\theta}$$

so Jefferys' prior has the form

$$p(\theta) \propto (J(\theta))^{\frac{1}{2}} = \left(\frac{1}{\theta}\right)^{\frac{1}{2}} = \theta^{-\frac{1}{2}}$$

and so $\theta \sim Beta\left(\frac{1}{2}, 1\right)$.
The $Gamma\left(\frac{1}{2}, 1\right)$ density looks to be a close match for small θ.



16. Comparing various $Beta(\alpha, \beta)$ prior distributions for $y|\theta \sim Binomial(m, \theta)$:

(a) $m = 6, y = 2, \dfrac{y}{m} = \dfrac{1}{3}$

$p(\theta)$	$E(\theta)$	$p(\theta\|y)$	$E(\theta\|y)$
$Beta(1,1)$	1/2	$Beta(3,5)$	3/8
$Beta(0.5,0.5)$	1/2	$Beta(2.5,4.5)$	5/14
$Beta(0,0)$	undefined	$Beta(2,4)$	1/3

(b) $m = 30, y = 10, \dfrac{y}{m} = \dfrac{1}{3}$

$p(\theta)$	$E(\theta)$	$p(\theta\|y)$	$E(\theta\|y)$
$Beta(1,1)$	1/2	$Beta(11,21)$	11/32
$Beta(0.5,0.5)$	1/2	$Beta(10.5,20.5)$	21/62
$Beta(0,0)$	undefined	$Beta(10,20)$	1/3

(c) $m = 90, y = 30, \dfrac{y}{m} = \dfrac{1}{3}$

$p(\theta)$	$E(\theta)$	$p(\theta\|y)$	$E(\theta\|y)$
$Beta(1,1)$	1/2	$Beta(31,61)$	31/92
$Beta(0.5,0.5)$	1/2	$Beta(30.5,60.5)$	61/182
$Beta(0,0)$	undefined	$Beta(30,60)$	1/3

In all cases, the $Beta(0,0)$ prior leads to a posterior mean that agrees with the maximum likelihood estimate that would be computed from the data alone. Increasing α and β pulls the posterior mean toward the prior mean. Observing more data results in the data dominating, with all three posterior means being close to 1/3 when 90 trials are observed.

17. Subjectivity refers to ideas that are shaped by the individual's experiences and can vary from person to person. Objectivity refers to ideas based on observable fact that everyone should be able to agree upon. I see these concepts relating to the steps of Statistical Inference in the following ways:

(a) Ask a question – Subjective The question that a particular researcher asks would depend on that researcher's previous experiences. A researcher observing starving polar bears who have lost their habitat due to thinning ice might ask if the mean weight of adult polar bears is below a certain values. A researcher who works with well-off polar bears in the vicinity of a particular Canadian island might ask if the mean weight of adult bears is above a certain value. Both researchers may be interested in collecting data from a larger population of bears.

(b) Design a study/experiment – Objective I see the design step as mostly objective since statisticians tend to agree the properties of a good design, and these conclusions are supported by mathematical results. Some subjectivity may appear through the reckless (mis)use of prior distributions.

(c) Collect data – Objective A solid design should include an unbiased sampling plan. I see data collection as objective as so long as the sampling scheme was clearly defined in the previous step. However, a poorly defined sampling plan, such using several ill-trained volunteers to stand at different locations around the MSU campus and conduct a survey, could introduce the volunteers' subjective opinions about who should be in the sample.

(d) Analyze data and make conclusions – Subjective The analysis component should be objective as far as the methods are concerned, since they are mainly mathematical. However, people often disagree on how to interpret results, how and when to trim terms from models, and what to do about outliers.

(e) Publish conclusions – Subjective Even in academia, unexpected or exciting conclusions are published disproportionately more often than results with that confirm existing ideas. This is exacerbated when the popular press focuses on new and surprising scientific results without considering the journey that the information took to get published, or the merits of the journal (or blog!) where it appeared. The results that a person is exposed to are determined by where that person seeks news and information.