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).
   -

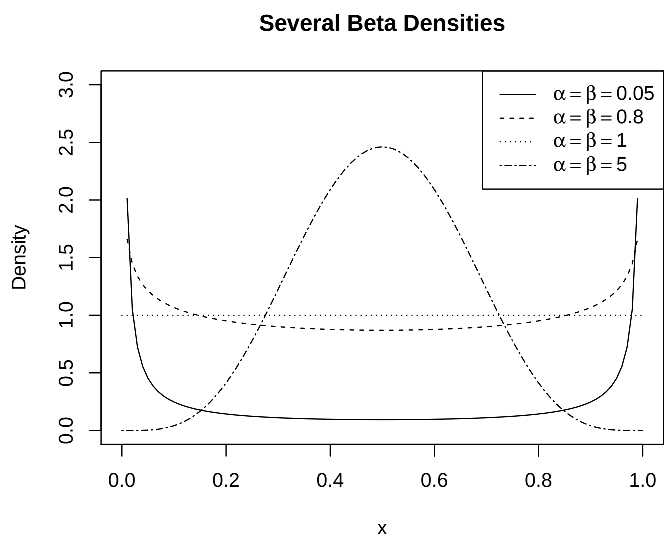
2.

3.

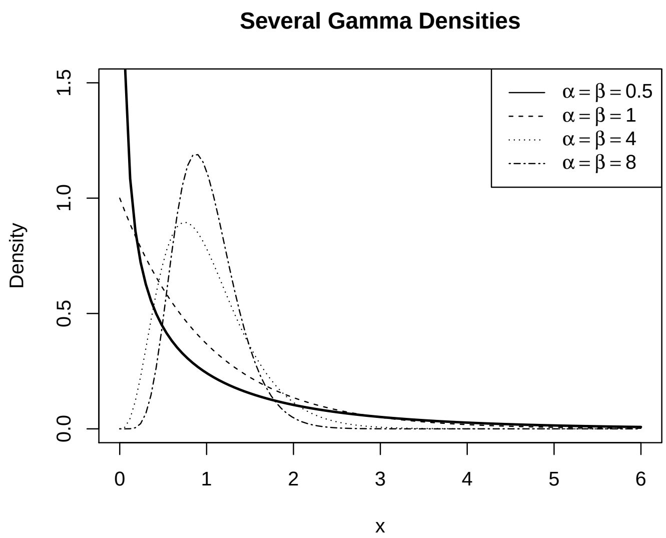
4.

5.

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,

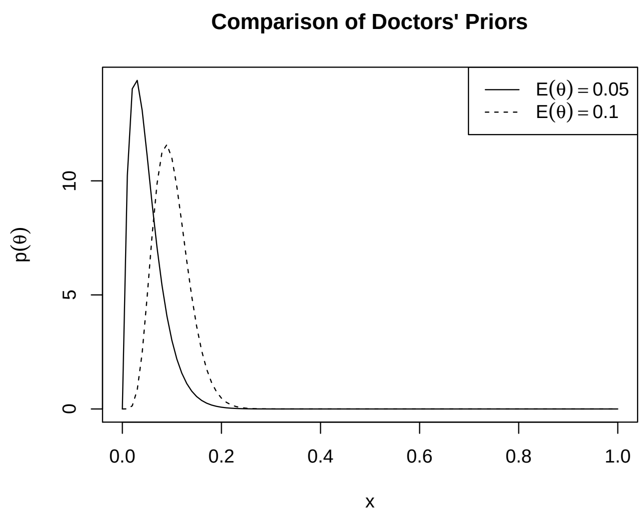
$$Pr(\text{infection}|\text{test }+) = \frac{(0.92)(0.1)}{(0.92)(0.1) + (0.08)(0.9)} = 0.561$$

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

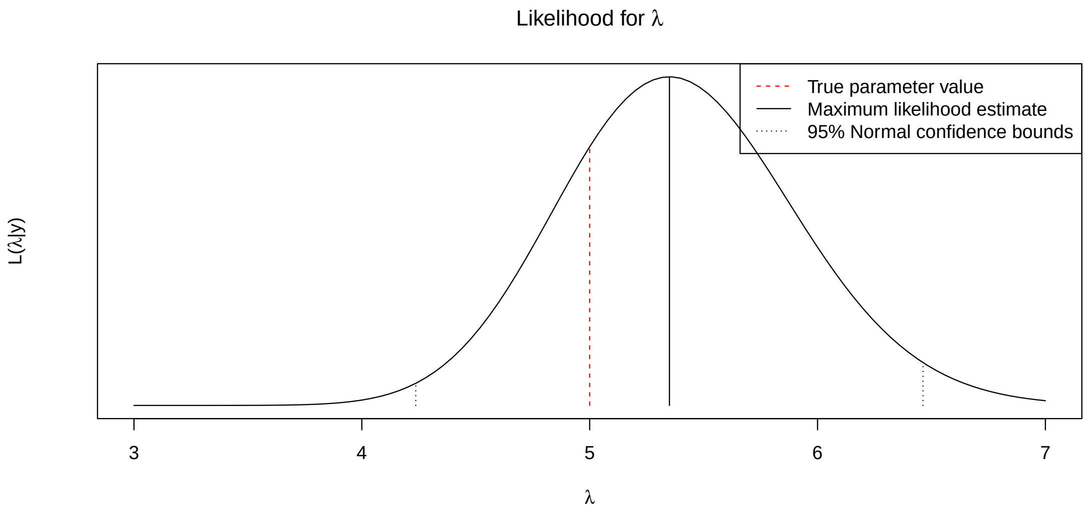
$$Pr(\theta < 0.5|y = 241945) = 1.$$

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

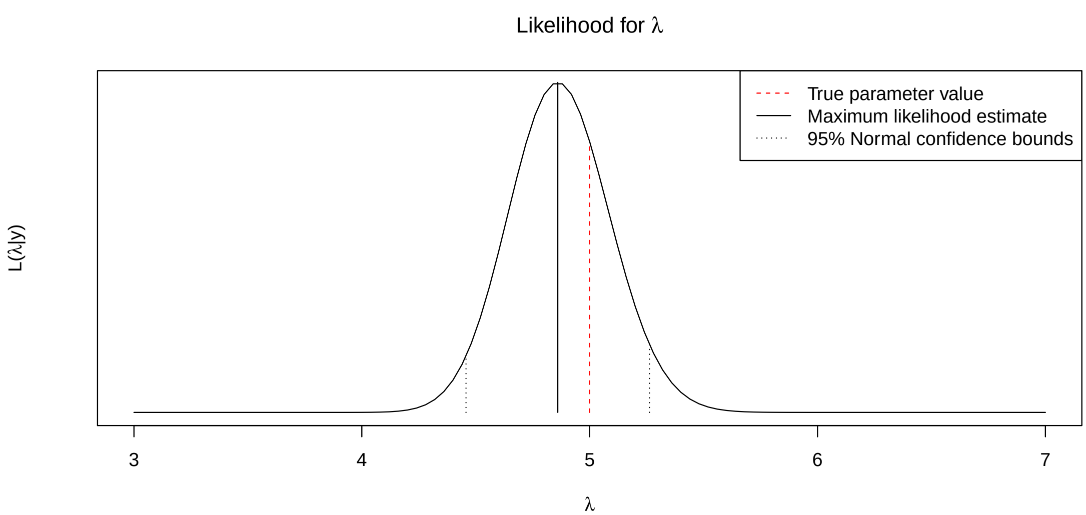
Maximum Likelihood Estimate: $\hat{\lambda} = \bar{y} = 5.35$, SE: $\dfrac{s}{\sqrt{n}} = 0.56789083$



(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

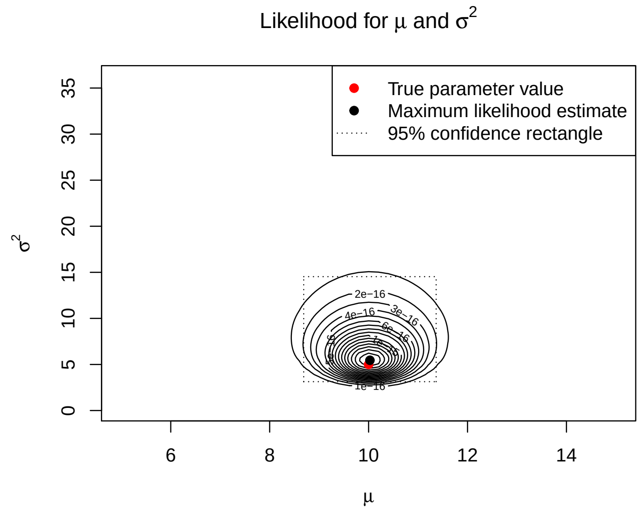
Maximum Likelihood Estimate: $\hat{\lambda} = \bar{y} = 4.86$, SE: $\dfrac{s}{\sqrt{n}} = 0.2054903$



(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: $\hat{\mu} = \bar{y} = 10.026587$, $SE(\bar{y}) = \dfrac{s}{\sqrt{n}} = 0.62402226$, $\hat{\sigma}^2 = \dfrac{\sum(y_i - \bar{y})^2}{n} = 5.451653$

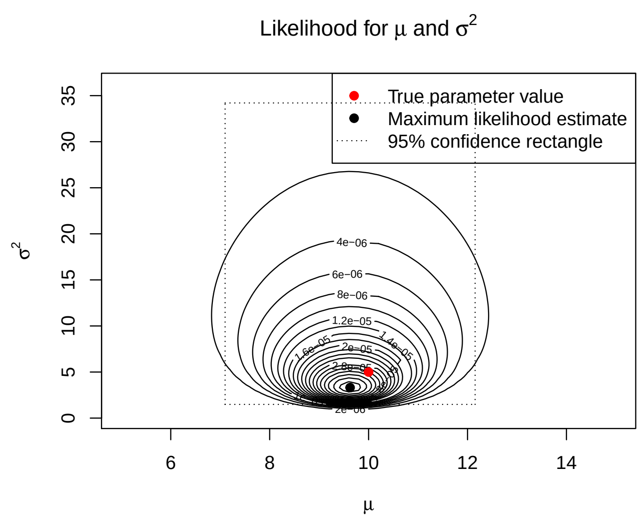


The confidence bounds are based on $\frac{\bar{y}-\mu}{s/sqrtn} \sim t_{14}$ and $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{14}$.

(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: $\hat{\mu} = \bar{y} = 9.6257655$, $SE(\bar{y}) = \frac{s}{\sqrt{n}} = 0.91021015$, $\hat{\sigma}^2 = \frac{\sum(y_i - \bar{y})^2}{n} = 3.3139301$

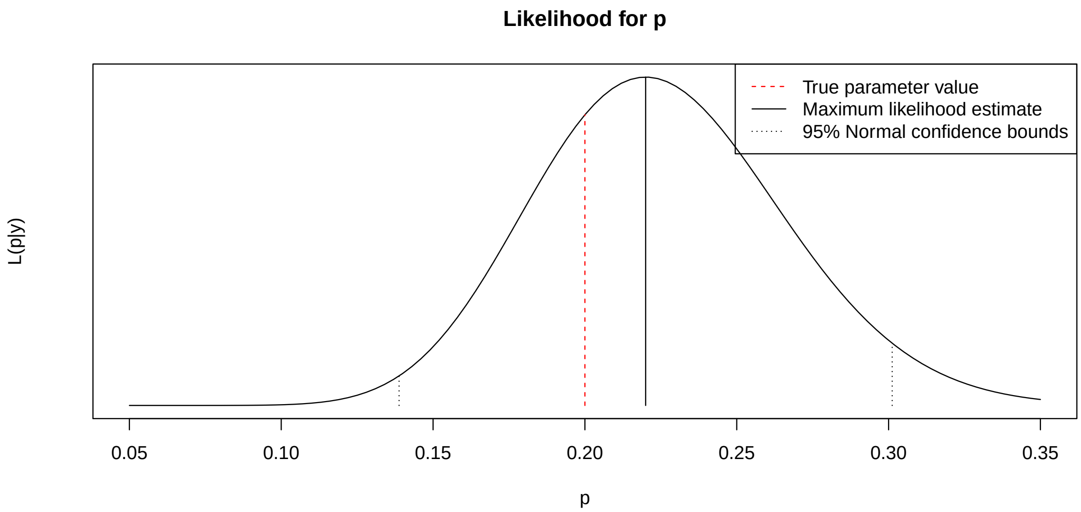


The confidence bounds are based on $\frac{\bar{y}-\mu}{s/\sqrt{n}} \sim t_4$ and $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_4$.

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

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

Maximum Likelihood Estimate: $\hat{p} = \frac{y}{m} = 0.22$, SE: $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.04142463$

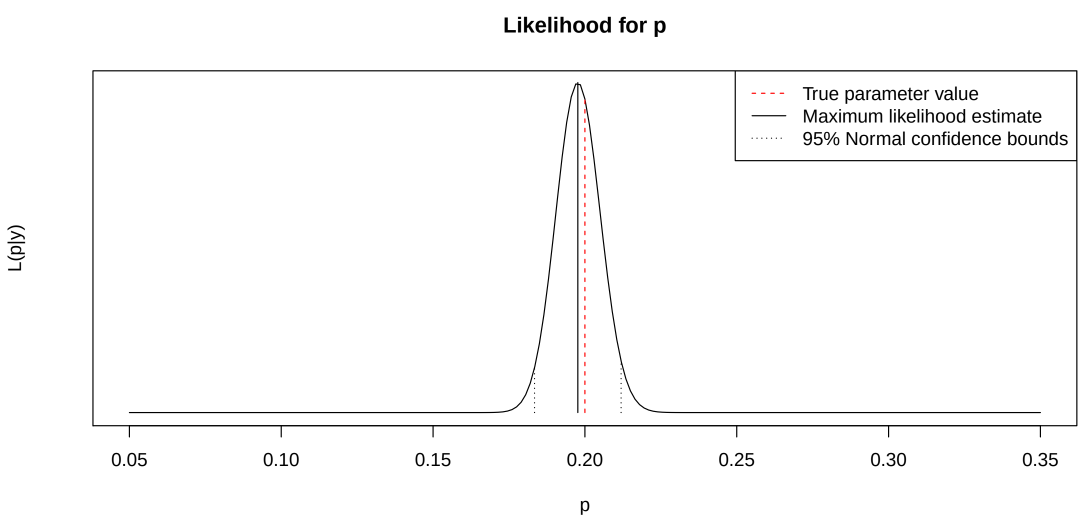


(f) 30 observations from $Binomial(m = 100, p = 0.2)$:

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: $\hat{p} = \dfrac{\sum y_i}{nm} = 0.19766667$, SE: $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{nm}} = 0.007270822$



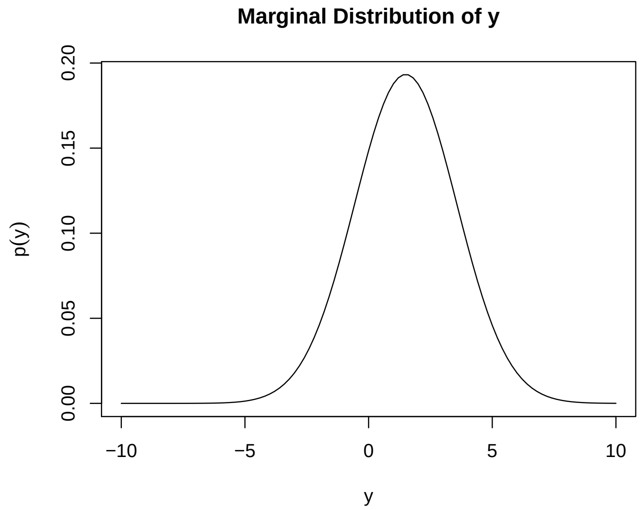
(g)

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, $-\frac{(y-\theta)^2}{2\sigma^2}$ decreases and the distribution is pulled towards y.

12.

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

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 $P_A(\text{Brazil wins the World Cup}) = \frac{1}{209}$.

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 qualifing 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 $P_B(\text{Brazil wins the World Cup}) = p > \frac{1}{32}$. 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 $E(\theta|y) = \frac{\alpha + y}{\alpha + \beta + n}$.

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

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

Therefore, if $\dfrac{\alpha}{\alpha+\beta} \neq \dfrac{y}{n}$ 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 $t = \dfrac{n}{2}$,

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

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

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

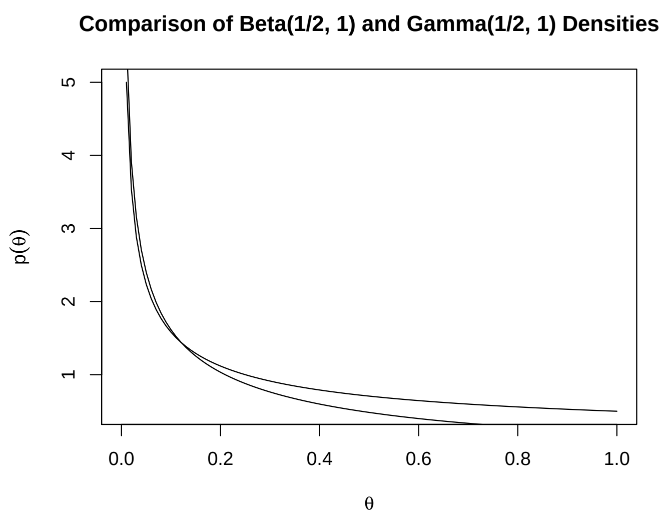
Then the Fisher information is

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

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 seems 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. Steps of Statistical Inference:

    (a) Ask a question
    (b) Design a study/experiment
    (c) Collect data
    (d) Analyze data
    (e) Make conclusions
    (f) Publish conclusions