1 Feb 2013

eBay/Google short course: Problem set 3 solutions

1. (a) The idea behind the change-of-variables formula is to work out the density of $\eta = g(\lambda) = \frac{1}{\lambda}$ by first expressing the CDF of η in terms of the CDF of λ and then differentiating. Here for $t > 0$

$$\begin{aligned} F_\eta(t) &= P(\eta \le t) = P\left(\frac{1}{\lambda} \le t\right) \stackrel{(i)}{=} P\left(\lambda \ge \frac{1}{t}\right) \\ &= 1 - P\left(\lambda < \frac{1}{t}\right) \stackrel{(ii)}{=} 1 - P\left(\lambda \le \frac{1}{t}\right) = 1 - F_\lambda\left(\frac{1}{t}\right), \end{aligned} \tag{1}$$

in which (i) relies on the fact that $\lambda > 0$ and (ii) is because λ is being treated as a continuous random variable (so that $P\left(\lambda = \frac{1}{t}\right) = 0$). Now differentiating the left- and right-most expressions in (1) gives

$$\frac{\partial}{\partial t} F_\eta(t) = p_\eta(t) = \frac{\partial}{\partial t}\left[1 - F_\lambda\left(\frac{1}{t}\right)\right] = p_\lambda\left(\frac{1}{t}\right)\, t^{-2}. \tag{2}$$

But for any argument $s > 0$, $p_\lambda(s) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{-(\alpha+1)} \exp\left(-\frac{\beta}{s}\right)$, so finally

$$p_\eta(t) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{t}\right)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\frac{1}{t}}\right)\, t^{-2} = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t), \tag{3}$$

which is recognizable as the $\Gamma(\alpha, \beta)$ density. Thus if $\lambda \sim \Gamma^{-1}(\alpha, \beta)$ then $\eta = \lambda^{-1} \sim \Gamma(\alpha, \beta)$, i.e., to go from one to the other you don't have to do anything complicated with the parameters α and β—they stay the same.

(b) In terms of λ the likelihood function is

$$l(\lambda|y) = c\,\lambda^{-n} \exp\left(-\frac{s}{\lambda}\right), \tag{4}$$

where $s = \sum_{i=1}^n y_i$; to get (4) in terms of η you just substitute $\lambda = \frac{1}{\eta}$, yielding

$$l(\eta|y) = c\,\eta^n \exp(-\eta s). \tag{5}$$

This is the $\Gamma(n+1, s)$ density, and the posterior is

$$p(\eta|y) = c\,p(\eta)\,l(\eta|y) = c\,\eta^{\alpha+n-1} e^{-(\beta+s)}, \tag{6}$$

i.e., $(\eta|y) \sim \Gamma(\alpha+n, \beta+s)$. With the values from Homework 2—$\alpha_0 = 8.25; \beta_0 = 32,625; n = 14; s = 70612$—the plot of the prior, likelihood, and posterior from `Maple` comes out as in Figure 1 below. To figure out the range over which to make this plot you can recall from Homework 2 that the posterior mean and SD for λ were 4858 and 1080, respectively, so that a 3–SD range for λ would run from about 1618 to 8098; therefore a 3–SD range for $\eta = \frac{1}{\lambda}$ should run from about $\frac{1}{8098} \doteq .00001$ to about $\frac{1}{1618} \doteq .0006$. I have used the slightly larger range $(0, .0007)$ just to be safe. The `Maple` code to produce Figure 1 is as follows:

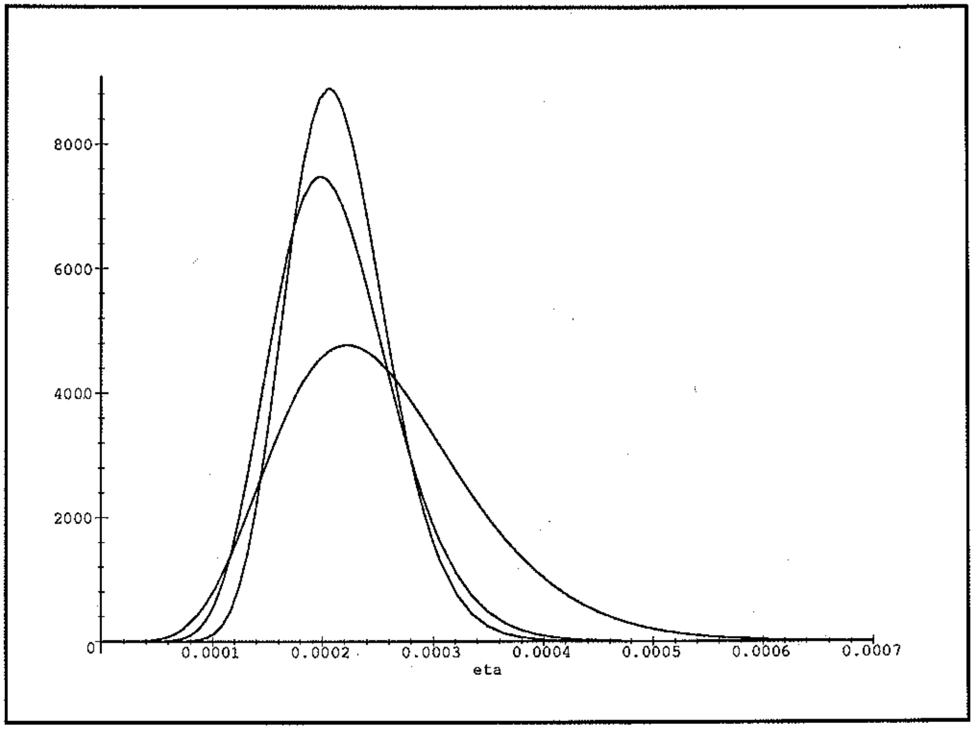

Figure 1: *Prior (shortest), likelihood (middle), and posterior (tallest) for* η *in the exponential failure model of Homework 1.*

```
rosalind 738> maple

    |\^/|     Maple V Release 5 (University of California, Santa Cruz)
._|\|   |/|_. Copyright (c) 1981-1997 by Waterloo Maple Inc. All rights
 \  MAPLE  /  reserved. Maple and Maple V are registered trademarks of
 <____ ____>  Waterloo Maple Inc.
      |       Type ? for help.

> gamma_density := ( eta, alpha, beta ) -> beta^alpha * eta^( alpha - 1 ) *
    exp( - beta * eta ) / GAMMA( alpha );

                                              alpha    (alpha - 1)
                                          beta      eta              exp(-beta eta)
gamma_density := (eta, alpha, beta) -> ---------------------------------------
                                                         GAMMA(alpha)

> alpha_0 := 8.25;

                                  alpha_0 := 8.25
```

```
> beta_0 := 32625;

                                beta_0 := 32625

> prior := gamma_density( eta, alpha_0, beta_0 );

                                          34    7.25
             prior := .2059342392 10   eta      exp(-32625 eta)

> n := 14;

                                    n := 14

> s := 70614;

                                  s := 70614

> likelihood := gamma_density( eta, n + 1, s );

likelihood :=

    10874917540669157487327304210061718755967424209596707663103000902416
    ---------------------------------------------------------------------
                                    175175

       14
    eta   exp(-70614 eta)

> posterior := gamma_density( eta, alpha_0 + n, beta_0 + s );

                                              92    21.25
           posterior := .3280441542 10    eta       exp(-103239 eta)

> plotsetup( x11 );

> plot( { prior, likelihood, posterior }, eta = 0 .. 0.0007, color = black );
```

(c) Using the facts that the $\Gamma(\alpha, \beta)$ distribution has mean $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$, and treating the likelihood literally as a $\Gamma(n+1, s)$ density, you get the following values for the prior, likelihood, and posterior means and SDs:

	Prior	Likelihood	Posterior
Mean	2.53e–04	2.12e–04	2.16e–04
SD	8.80e–05	5.48e–05	4.57e–05

If you try to literally solve the equation which would attempt to express the posterior mean as a weighted average of the prior and likelihood means,

$$\frac{\alpha_0 + n}{\beta_0 + s} = w\left(\frac{\alpha_0}{\beta_0}\right) + (1 - w)\left(\frac{n+1}{s}\right), \tag{7}$$

for the weight w, you get

$$w = \left(\frac{\beta_0}{\beta_0 + s}\right)\left[1 - \frac{s}{s\alpha_0 - (n+1)\beta_0}\right]. \tag{8}$$

But to be a proper weighted average this expression has to be between 0 and 1, and this will only be true if

$$s\alpha_0 \geq (n+1)\beta_0 \quad \longleftrightarrow \quad \frac{\alpha_0}{\beta_0} \geq \frac{n+1}{s}, \tag{9}$$

i.e., if the prior mean is at least as large as the likelihood mean. That does happen, coincidentally, to be true for this data set and prior but would not be true in general, so you can't always express the posterior mean as a weighted average of the prior and likelihood means when the quantity of interest is $\eta = \frac{1}{\lambda}$. But if you make an $O\left(\frac{1}{n}\right)$ approximation by replacing the likelihood mean $\frac{n+1}{s}$ by $\frac{n}{s}$ (in effect this treats the likelihood distribution as $\Gamma(n, s)$ instead of $\Gamma(n+1, s)$), the following approximate expression results:

$$\frac{\alpha_0 + n}{\beta_0 + s} \doteq \frac{\beta_0\left(\frac{\alpha_0}{\beta_0}\right) + s\left(\frac{n}{s}\right)}{\beta_0 + s}. \tag{10}$$

Within the scope of this approximation there's an analogy between $\{\alpha_0$ and $n\}$ on the one hand and $\{\beta_0$ and $s\}$ on the other hand, so I guess you could say that α acts like the prior sample size and β like the *sum* of the values in the data set that the prior is equivalent to.

2. (a) Since, as the problem says, the only way all of the y_i can be $\leq \theta$ is if m, the largest y_i, is $\leq \theta$,

$$\begin{aligned} l(\theta|y) &= \prod_{i=1}^{n} p(y_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I\left(0 \leq y_i \leq \theta\right) \\ &= \theta^{-n} I(0 \leq y_i \leq \theta \text{ for all } i = 1, \ldots, n) = \theta^{-n} I(m \leq \theta). \end{aligned} \tag{11}$$

m is thus sufficient for θ in this model because the likelihood function depends on the data vector $y = (y_1, \ldots, y_n)$ only through m.

(b) A rough sketch of the likelihood function looks like

and this function is clearly maximized at $\theta = \hat{\theta}_{\text{MLE}} = m$. The maximum of the likelihood or log likelihood function occurs at a point of sharp discontinuity, so derivative-based methods for finding extreme points fail in this case.

(c) Rewriting equation (12) of the homework assignment as

$$p(\theta) = \alpha\, \beta^{\alpha}\, \theta^{-(\alpha+1)} I(\theta \geq \beta), \qquad (12)$$

and ignoring the constant $\alpha\, \beta^{\alpha}$, the likelihood function is evidently a constant multiple of the Pareto$(n-1, m)$ distribution. If the prior is Pareto(α, β) and the likelihood is Pareto$(n-1, m)$ then the posterior must be

$$\begin{aligned} p(\theta|y) &= c\, \alpha\, \beta^{\alpha}\, \theta^{-(\alpha+1)} I(\theta \geq \beta)\, \theta^{-n} I(\theta \geq m) \\ &= c\, \theta^{-(\alpha+n+1)} I[\theta \geq \max(\beta, m)], \end{aligned} \qquad (13)$$

which is recognizable as the Pareto $[\alpha + n, \max(\beta, m)]$ distribution, i.e., the Pareto prior is conjugate to the Uniform$(0, \theta)$ likelihood.

(d) The plot, and the `Maple` code to generate it, are below.

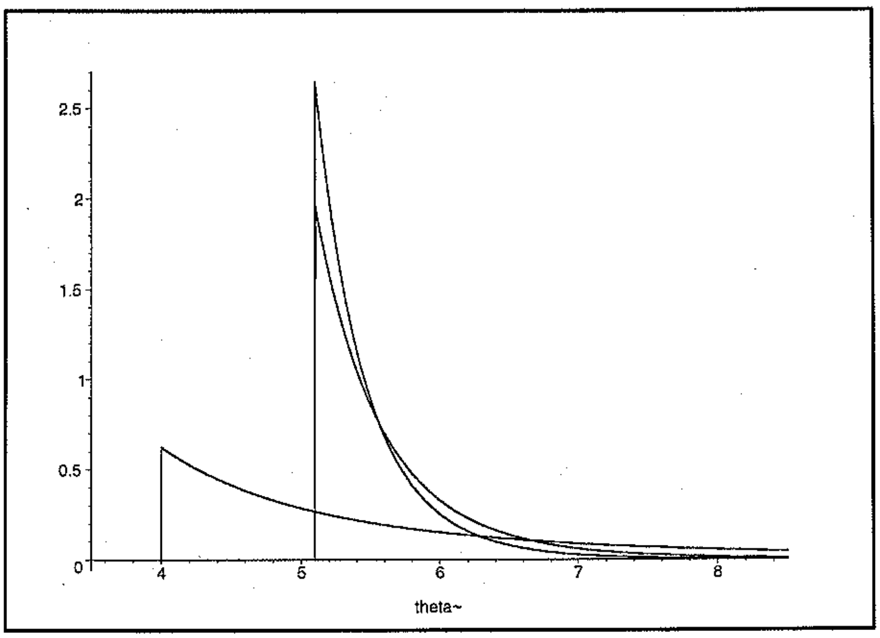

Figure 2: *Prior (shortest), likelihood (middle), and posterior (tallest) for θ with the paleobotany data.*

```
rosalind 764> maple

    |\^/|     Maple V Release 5 (University of California, Santa Cruz)
._|\|   |/|_. Copyright (c) 1981-1997 by Waterloo Maple Inc. All rights
 \  MAPLE  /  reserved. Maple and Maple V are registered trademarks of
 <____ ____>  Waterloo Maple Inc.
      |       Type ? for help.
```

```
> assume( alpha > 0, beta > 0, theta > 0 );

> pareto := ( alpha, beta, theta ) -> piecewise(

  theta < beta, 0,
  theta >= beta, alpha * beta^alpha * theta^( - ( alpha + 1 ) )

  );

pareto := (alpha, beta, theta) -> piecewise(theta < beta, 0, beta <= theta,

                 alpha       (-alpha - 1)
    alpha beta         theta                )


> plotsetup( x11 );

> plot( { pareto( 2.5, 4, theta ), pareto( 10.0, 5.1, theta ),
    pareto( 13.5, 5.1, theta ) }, theta = 3.5 .. 10.0, color = black );
```

The prior is evidently the distribution which begins at $\beta_0 = 4$. The posterior is always at least as peaked as the likelihood (to reflect the combining of prior and data information), so it's the tallest curve, and the likelihood is in between. Here the posterior is noticeably more concentrated near 5.1 than the likelihood is, reflecting the substantial contribution of the prior information.

(e) Plugging the indicated values into the mean and variance formulas for the Pareto yields Table 1 below.

Table 1. *Summaries of a Bayesian analysis of the paleobotany data.*

Summary	Prior	Likelihood	Posterior
Mean	6.67	5.67	5.51
SD	5.96	0.63	0.44

The prior mean, for example, is $\frac{\alpha\beta}{\alpha-1} = \frac{(2.5)(4)}{1.5} = 6.67$, and the likelihood SD is $\sqrt{\frac{(n-1)m^2}{(n-2)^2(n-3)}} = \sqrt{\frac{10(5.1)^2}{9^2 \cdot 8}} = 0.63$. Here, while it is true that the posterior SD is smaller than either the prior or likelihood SDs, it is—unusually, and interestingly—not true that the posterior mean is a weighted average of the prior and likelihood means (the former is smaller than either of the latter). In fact, it's the posterior *mode* that has to lie somewhere between the prior and likelihood modes (inclusive) in this model, not the mean.

(f) The posterior is Pareto$[\alpha + n, \max(\beta, m)]$ and so has variance

$$V(\theta|y) = \frac{(\alpha + n)\left[\max(\beta, m)\right]^2}{(\alpha + n - 1)^2(\alpha + n - 2)}. \tag{14}$$

The unusual thing about this expression is that it is $O\left(\frac{1}{n^2}\right)$, i.e., it goes to zero at rate $\frac{1}{n^2}$, whereas in every other example in this course the posterior variance (for parameter estimation, not prediction) decreases like $\frac{1}{n}$. In all of our earlier examples the parameters either tracked the location or scale of the distribution; here θ is a *range-restricting* parameter, and the moral of the story is that learning about such parameters takes place much faster than with location or scale parameters. This is reflected in Table 1 above: look at how much smaller the posterior SD is than the prior SD (even with a sample size of only 10).

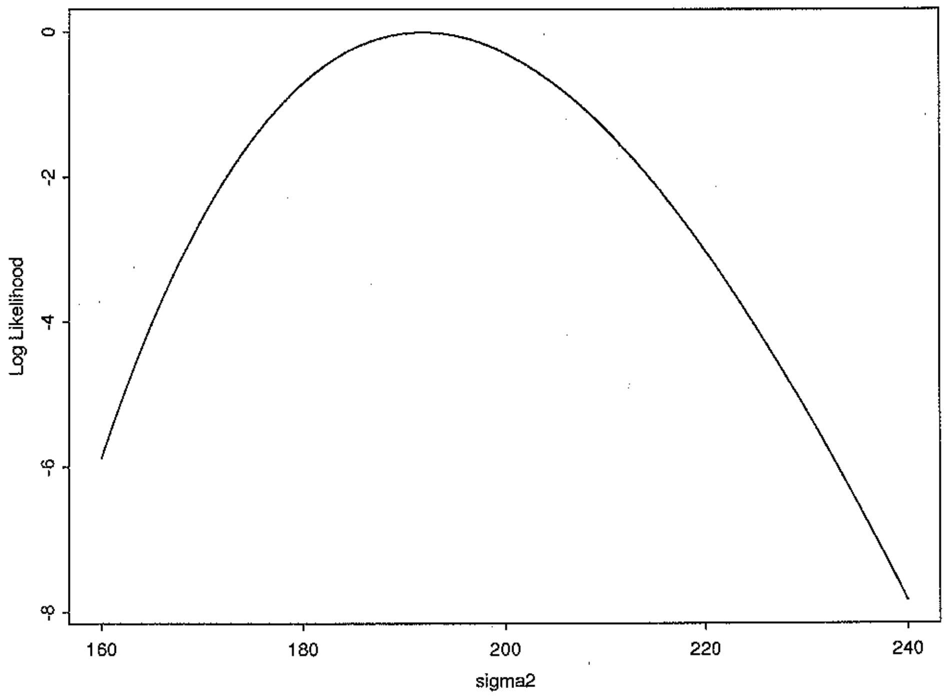

Figure 3: *Log likelihood in the* $N(0, \sigma^2)$ *model for the football point spread data.*

3. (a) The likelihood and log likelihood functions for σ^2 are

$$\begin{aligned} l(\sigma^2|d) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{d_i^2}{2\sigma^2}\right) = c\left(\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{SSQ}{2\sigma^2}\right) \quad \text{and} \\ ll(\sigma^2|d) &= c - \frac{n}{2}\log(\sigma^2) - \frac{SSQ}{2\sigma^2}, \end{aligned} \tag{15}$$

where $SSQ = \sum_{i=1}^{n} d_i^2$, or any nontrivial function of it like $\hat{\sigma}^2 = \frac{1}{n} SSQ$, is evidently sufficient for σ^2. Differentiating the log likelihood with respect to σ^2 and setting to 0 gives

$$\frac{\partial}{\partial \sigma^2} ll(\sigma^2|d) = -\frac{n}{2\sigma^2} + \frac{SSQ}{2\left(\sigma^2\right)^2} = 0 \quad \text{iff} \quad \sigma^2 = \hat{\sigma}^2_{MLE} = \frac{SSQ}{n}. \tag{16}$$

The log likelihood function is plotted in Figure 3; its shape guarantees that the critical point found in (16) is the global max. The slight skewness evident in this plot corresponds to a mild long right-hand tail in the sampling distribution of $\hat{\sigma}^2$, which is to be expected for variance estimates even with large n.

(b) Multiplying prior $p(\sigma^2) = c_0 \, (\sigma^2)^{-1}$ and likelihood $l(\sigma^2|d) = c \, (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{SSQ}{2\sigma^2}\right)$ gives

$$p(\sigma^2|d) = c \, \left(\sigma^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left(-\frac{SSQ}{2\sigma^2}\right), \qquad (17)$$

which is recognizable from equation (13) of the homework assignment as the χ^{-2} distribution with parameters n and $\frac{1}{n}SSQ$, i.e., $\chi^{-2}(n, \hat{\sigma}^2)$. The prior is evidently the nearly-constant curve (the small dotted lines) in Figure 4 below, and it will act to slightly pull the likelihood back toward 0, so the likelihood and posterior densities must be the large dotted and solid lines, respectively. It is evident that the "noninformative" prior has been successful in forcing the posterior to virtually coincide with the likelihood with these data.

As I mentioned in class, `Maple` is surprisingly good at working with enormously big and enormously small numbers, but its abilities in this regard are limited, and this problem stretches them beyond the breaking point when you try to do the calculations for the density plots on the raw density scale (for example, on that scale you end up asking `Maple` to multiply numbers like 192^{335} times 200^{-335}, and 192^{335} is on the order of 10^{766} (**NB** if you've ever heard of a *googol*, that term was invented to describe an astoundingly big number, 10^{100}; these numbers are bigger). Here's some quite straightforward code in `R` to make Figure 4 below; this relies on recognizing that the likelihood is equivalent to a positive constant times the $\chi^{-2}(n-2, \frac{n}{n-2}\hat{\sigma}^2)$ distribution. The main point of the code is to evaluate the density on the log scale (using the `lngamma` function) and exponentiate at the end, to avoid working with fantastically small and fantastically large numbers multiplied together.

```
> sichi2 <- function( theta, nu, s2 ) {

    return( exp( ( nu / 2 ) * log( nu / 2 ) +
      ( nu / 2 ) * log( s2 ) - ( nu / 2 + 1 ) * log( theta ) - nu * s2 /
      ( 2 * theta ) - lgamma( nu / 2 ) ) )

  }

> n <- 672

> s2 <- 191.8

> x11( )

> theta <- seq( 160, 240, length = 500 )

> plot( theta, sichi2( theta, n, s2 ), type = 'l', xlab = 'sigma2',
    ylab = 'Density' )

> lines( theta, sichi2( theta, n - 2, n * s2 / ( n - 2 ) ), lty = 2 )
```

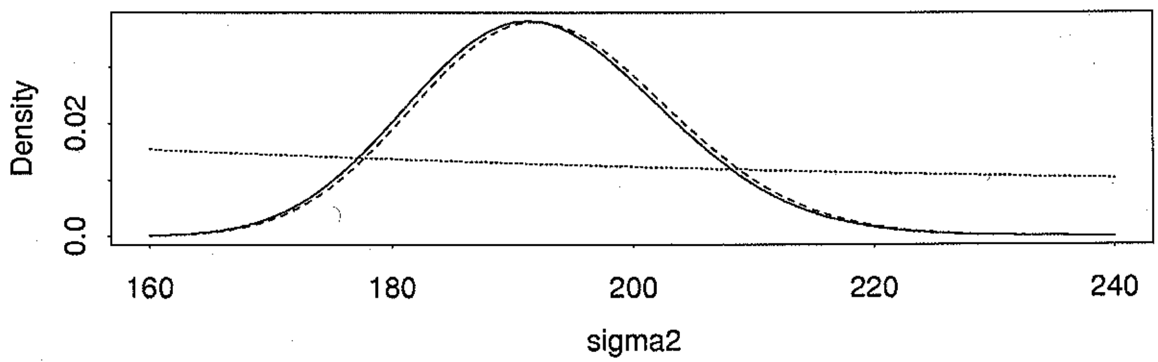

Figure 4: *Prior, likelihood, and posterior densities with the football point spread data.*

```
> lines( theta, 2.5 / theta, lty = 3 )
```

4. (a) This process is clearly Markovian, because you only need to know how much money Y_t the gambler has at any given time t to determine the probability distribution for where her fortune will be at time $(t+1)$, i.e., the past history of the process is irrelevant in simulating her next move. Assuming that none of the following values goes above N (I'm going to omit all monetary units in the interests of brevity) or below 0 (when her fortune reaches N or 0 the game stops), here are the possible states as time unfolds: at $t = 0$ her fortune Y_0 has to be M; at $t = 1, Y_1$ can only be $(M-1)$ or $(M+1)$; at $t = 2, Y_2$ can only be $(M-2), M$, or $(M+2)$; at $t = 3, Y_3$ can only be $(M-3), (M-1), (M+1)$, or $(M+3)$; and so on. Every time the chain tries to go below 0 or above N it has to try to pass through 0 or N to do so, and states 0 and N are *absorbing* (once you reach them you can't leave them), so clearly the only possible states for the chain are $\{0, 1, \ldots, N\}$. Notice that this Markov chain is not aperiodic—it has period 2 (the amount of time it takes to get back to any state, having started there, is a multiple of 2).

(b) I'll sketch the mathematical solution to this problem (based on the treatment in the 1970 Holden-Day book by SM Ross called *Applied Probability Models with Optimization Applications*); if you worked on it from a simulation point of view you probably got results that are similar to those described here. Before going any further, let's think about the qualitative behavior that P should satisfy as a function of M, N, and p.

(1) For fixed N and p, P should increase as M increases, because the more money you have to start with the easier it is to break the bank (that's shorthand for reaching N and causing the casino to admit defeat);

(2) For fixed M and p, P should decrease as N increases, since it's harder to break the bank if the casino has more money to begin with; and

(3) For fixed M and N, P should increase as p increases, because it's easier to win the overall game if your winning probability on any single play is higher.

For any states i and j in a finite-state-space Markov chain, let f_{ij}^t be the probability that, starting in i, the first transition into j occurs at time t. Then $f_{ij} = \sum_{t=1}^{\infty} f_{ij}^t$ is the probability of ever making a transition into j, given that the chain started in i. Any given state i is called *recurrent* if $f_{ii} = 1$, and *transient* otherwise. Thus i is recurrent iff with probability 1 the process will return to i if it started there. Because states 0 and N are absorbing, they're both (trivially) recurrent, and moreover all the other states are transient (assuming that the probabilities $(1-p)$ and p of moving left and right on any given play are both strictly between 0 and 1, if you start anywhere from 1 to $(N-1)$ there is a positive probability you will be absorbed at 0 or N before you get back to where you started). Denoting (as we did in class) by P_{ij} the probability of moving from i to j in one iteration, and letting T stand for the set of transient states for the process, it's a basic fact from Markov chain theory that

If j is recurrent, then the set of probabilities $\{f_{ij}, i \in T\}$ satisfies the relation

$$f_{ij} = \sum_{k \in T} P_{ik} f_{kj} + \sum_{k \in R} P_{ik} \quad \text{for any } i \in T, \tag{18}$$

where R denotes the set of states *communicating with* j.

(Letting P_{ij}^t be the probability of moving from i to j in t iterations, state j is *accessible* from state i if for some $t \geq 0, P_{ij}^t > 0$; two states i and j that are accessible to each other are said to *communicate*.) From the result in the box above, the f_{ij} for the gambler's ruin problem must satisfy the following equations:

$$\begin{aligned} f_{0N} &= 0 \\ f_{iN} &= p\, f_{i+1,N} + (1-p)\, f_{i-1,N} \quad \text{for } i = 1, \ldots, (N-1) \\ f_{NN} &= 1. \end{aligned} \tag{19}$$

Rearrange these equations and write f_i for f_{iN} to get

$$f_{i+1} - f_i = \left(\frac{1-p}{p}\right)(f_i - f_{i-1}) \quad \text{for } i = 1, \ldots, (N-1). \tag{20}$$

But this defines a recurrence relation among the f_i:

$$\begin{aligned} f_2 - f_1 &= \left(\frac{1-p}{p}\right) f_1 \\ f_3 - f_2 &= \left(\frac{1-p}{p}\right)(f_2 - f_1) = \left(\frac{1-p}{p}\right)^2 f_1 \\ \vdots &= \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ 1 - f_{N-1} &= \left(\frac{1-p}{p}\right)(f_{N-1} - f_{N-2}) = \left(\frac{1-p}{p}\right)^{N-1} f_1 \end{aligned} \tag{21}$$

Add these equations to get

$$f_i - f_1 = f_1 \left[\left(\frac{1-p}{p}\right) + \left(\frac{1-p}{p}\right)^2 + \ldots + \left(\frac{1-p}{p}\right)^{i-1} \right] \quad \text{for } i > 1, \tag{22}$$

which can be simplified (remembering how geometric series work) to yield

$$f_i = \left\{ \begin{array}{cc} \frac{1-\left(\frac{1-p}{p}\right)^i}{1-\left(\frac{1-p}{p}\right)} f_1 & \text{if } \frac{1-p}{p} \neq 1 \\ i\, f_1 & \text{if } \frac{1-p}{p} = 1 \end{array} \right\} \quad \text{for } i > 1. \tag{23}$$

Finally use the fact that $f_N = 1$ and simplify further to obtain that for $i = 1, \ldots, N$

$$f_{iN} = \left\{ \begin{array}{cc} \frac{1-\left(\frac{1-p}{p}\right)^i}{1-\left(\frac{1-p}{p}\right)^N} & \text{if } p \neq \frac{1}{2} \\ \frac{i}{N} & \text{if } p = \frac{1}{2} \end{array} \right\}. \tag{24}$$

Since state 0 is absorbing, substituting $i = M$ in the expression for f_{iN} in (24) gives the probability we wanted, namely the chance P that the gambler will break the bank (reach N) before going broke (reach 0):

$$P = \left\{ \begin{array}{cc} \frac{1-\left(\frac{1-p}{p}\right)^M}{1-\left(\frac{1-p}{p}\right)^N} & \text{if } p \neq \frac{1}{2} \\ \frac{M}{N} & \text{if } p = \frac{1}{2} \end{array} \right\}. \tag{25}$$

Perhaps the first thing to note about (25) is that it's continuous in p (i.e., the limit of the upper expression as $p \to \frac{1}{2}$ is the lower expression, as you can verify with an application of l'Hospital's rule or a session with `Maple`). The next thing to note (by inspection) is that (25) satisfies all three of the qualitative behaviors mentioned on the bottom of page 9. I recommend an extended plotting session with `Maple` to discover other properties of (25); since it depends on three quantities it's hard to summarize its behavior neatly (to take a practical example of direct interest to the gambler, can you succinctly specify the regions in (M, N, p) space such that $P > \frac{1}{2}$?).

(c) As $N \to \infty$ with the gambler's initial capital M held fixed, things simplify: for $p = \frac{1}{2}$, evidently $P \to 0$ in (25); for $p < \frac{1}{2}$, $\frac{1-p}{p} > 1$ and it's better to think of (25) as

$$P = \frac{\left(\frac{1-p}{p}\right)^M - 1}{\left(\frac{1-p}{p}\right)^N - 1}. \tag{26}$$

As $N \to \infty$ the denominator of (26) becomes indefinitely large with the numerator fixed, so in this case it's also true (how could it be otherwise?) that $P \to 0$. If $p > \frac{1}{2}$, however, there's some hope for the gambler; in that case $\frac{1-p}{p} < 1$ and the denominator of the top expression in (25) goes to 1 as $N \to \infty$, so in the limit $P \to 1 - \left(\frac{1-p}{p}\right)^M$ (we would interpret this as the probability of the gambler's fortune increasing indefinitely). As $p \to 1$ this expression goes to 1, and as M increases it also tends to 1, both of which make good qualitative sense. The most interesting of all these cases is $p = \frac{1}{2}$: even if she starts with a huge value of M, if the game is fair and the casino has an infinite initial supply of money she will eventually go broke (because her goal is infinitely far to the right of where she started on the number line).