Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam January 8, 2024, 8:30 - 12:30
Examiner: Petter Mostad will be available by phone 031-772-3579
and will visit the exam at 9:30 and 11:30.
Allowed aids: Chalmers-approved calculator
Total number of points: 30. At least 12 points are needed to pass.
See appendix for some information about some probability distributions.
All answers need to be explicitly computed or explicitly argued for.

1. (6 points) Consider a Branching process with offspring distribution X, where $X = Y - 1$ and $Y \sim \text{Geometric}(p)$, with $0 < p < 1$.
   (a) Compute the expectation μ and the variance σ^2 of the offspring distribution.
   (b) Find the probability of extinction as a function of p.
   (c) Compute the variance of Z_2, the size of the population after 2 generations, by conditioning on Z_1, the size of the population after 1 generation.

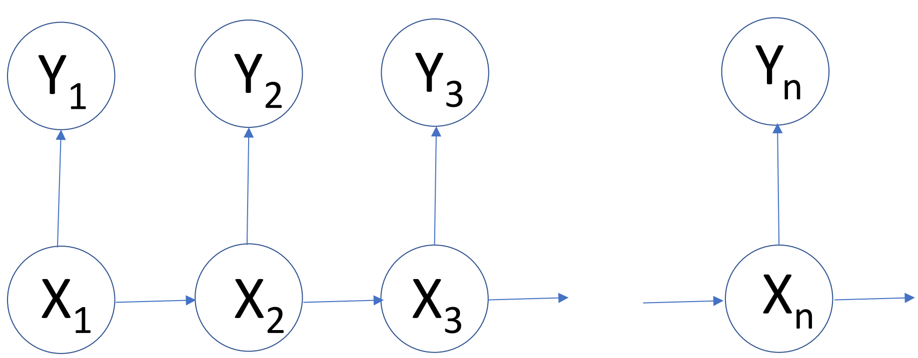

Figure 1: The hidden Markov model used in question 2.

2. (6 points) Consider the Hidden Markov Model illustrated in Figure 1. The random variables $X_1, X_2, \dots$ have possible values 1,2,3, or 4. The random variables $Y_1, Y_2, \dots$ have

non-negative real values[1]. In fact, the relationship between X_i and Y_i is described with

$$Y_i \mid X_i \sim \text{Exponential}(\lambda X_i).$$

where λ is an unknown parameter. We use the prior $\pi(\lambda) \propto 1/\lambda$. You are given data containing observations $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$ and $y_1, \ldots, y_n$ of $Y_1, \ldots, Y_n$. Example data is given in Table 1 for $n = 10$.

x	2	2	1	3	4	3	2	1	2	4
y	2.21	6.26	0.72	0.33	0.31	11.33	0.05	10.14	1.50	1.43

Table 1: Data for question 2b.

(a) Describe a possible prior for the transition matrix P of the Markov chain $X_1, X_2, \ldots$, which contains the information that we have $X_{i+1} - X_i \geq -1$ for any i, i.e., in the chain of x values, any jump downwards is with at most one step.

(b) Using the data in Table 1 and the prior you described in (a), compute the probability distribution for the possible values of X_{11}.

(c) What is the posterior for λ given data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$? (Find a formula, do not use the specific data of Table 1).

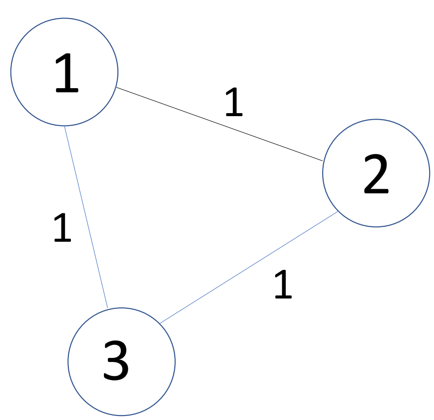

Figure 2: The undirected weighted graph used in question 3.

[1] In the original exam, the indexation started at 0 instead of at 1. This has been changed to give a more coherent notation.

3. (6 points) Consider the Metropolis Hastings algorithm applied to the state space with the three elements $\{1, 2, 3\}$, with target distribution $\pi(1) = 0.8$, $\pi(2) = 0.1$, and $\pi(3) = 0.1$, and proposal function given by the random walk on the undirected weighted graph shown in Figure 2.

   (a) Compute the terms of the transition matrix T for the Metropolis Hastings Markov chain defined above.

   (b) Is this Markov chain time reversible? Give an argument for your answer.

   (c) Find a representation of the Metropolis Hastings Markov chain as a random walk on a weighted undirected graph: Draw the graph and its weights.

4. (6 points) A continuous-time Markov chain has 4 possible states: We call them 1,2,3, and 4. From state 1, it moves to state 2 with rate 2. From state 2, it moves to state 1 with rate 2 and to state 3 with rate 1. When in state 3, it moves out of that state with rate 3, and into each of the other states with equal probability. When in state 4, it moves to state 3 with rate 1.

   Below, numerical answers may, if you like, be expressed as an equation involving numerical vectors and numerical matrices: You do not need to include the final numerical answer for such expressions.

   (a) Draw a transition rate graph for the chain.

   (b) Write down its generator matrix Q, and $\tilde{P}$, the transition matrix for the embedded chain.

   (c) Argue why the continuous-time Markov chain has a limiting distribution, and find this limiting distribution.

   (d) Assume the chain starts in state 2. Compute the expected time until the first time it reaches either state 3 or state 4.

5. (6 points) A cog-wheel in a machine has $n = 14$ cogs (or "spikes") and turns forward one cog at a time. The time it takes for it to move one cog forward is Exponentially distributed with parameter $\lambda = 0.3$.

   Below, you may give your answers as a numerical value, or R code for computing the numerical answer, or, if you don't remember the R syntax, as an equivalent mathematical expression for computing the numerical answer.

   (a) What is the probability that it does not move at all in the time interval $[0, 2]$?

   (b) Given the information that at time 20 it has moved exactly a full turn, what is the probability that it did not move at all in the time interval $[0, 2]$?

   (c) What is the probability that the wheel has completed 3 full turns before the time $t = 100$?

Appendix: Some probability distributions

The Beta distribution

If $x \in [0,1]$ has a Beta distribution with parameters with $\alpha > 0$ and $\beta > 0$ then the density is

$$\pi(x \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

We write $x \mid \alpha, \beta \sim \text{Beta}(\alpha, \beta)$ and $\pi(x \mid \alpha, \beta) = \text{Beta}(x; \alpha, \beta)$.

The Beta-Binomial distribution

If $x \in \{0, 1, 2, \ldots, n\}$ has a Beta-Binomial distribution, with n a positive integer and parameters $\alpha > 0$ and $\beta > 0$, then the probability mass function is

$$\pi(x \mid n, \alpha, \beta) = \binom{n}{x} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}.$$

We write $x \mid n, \alpha, \beta \sim \text{Beta-Binomial}(n, \alpha, \beta)$ and $\pi(x \mid n, \alpha, \beta) = \text{Beta-Binomial}(x; n, \alpha, \beta)$.

The Binomial distribution

If $x \in \{0, 1, 2, \ldots, n\}$ has a Binomial distribution, with n a positive integer and $0 \leq p \leq 1$, then the probability mass function is

$$\pi(x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

We write $x \mid n, p \sim \text{Binomial}(n, p)$ and $\pi(x \mid n, p) = \text{Binomial}(x; n, p)$.

The Dirichlet distribution

If $x = (x_1, x_2, \ldots, x_n)$ has a Dirichlet distribution, with $x_i \geq 0$ and $\sum_{i=1}^{n} x_i = 1$ and with parameters $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_1 > 0, \ldots, \alpha_n > 0$, then the density function is

$$\pi(x \mid \alpha) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_n)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_n^{\alpha_n - 1}.$$

We write $x \mid \alpha \sim \text{Dirichlet}(\alpha)$ and $\pi(x \mid \alpha) = \text{Dirichlet}(x; \alpha)$.

The Exponential distribution

If $x \geq 0$ has an Exponential distribution with parameter $\lambda > 0$, then the density is

$$\pi(x \mid \lambda) = \lambda \exp(-\lambda x)$$

We write $x \mid \lambda \sim \text{Exponential}(\lambda)$ and $\pi(x \mid \lambda) = \text{Exponential}(x; \lambda)$. The expectation is $1/\lambda$ and the variance is $1/\lambda^2$.

The Gamma distribution

If $x > 0$ has a Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ then the density is

$$\pi(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

We write $x \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta)$ and $\pi(x \mid \alpha, \beta) = \text{Gamma}(x; \alpha, \beta)$.

The Geometric distribution

If $x \in \{1, 2, 3, \dots\}$ has a Geometric distribution with parameter $p \in (0, 1)$, the probability mass function is

$$\pi(x \mid p) = p(1-p)^{x-1}$$

We write $x \mid p \sim \text{Geometric}(p)$ and $\pi(x \mid p) = \text{Geometric}(x; p)$. The expectation is $1/p$ and the variance $(1-p)/p^2$.

The Negative Binomial distribution

A stochastic variable x taking on as possible values any nonnegative integer has a Negative Binomial distribution if its probability mass function is given by

$$\pi(x \mid r, p) = \binom{x+r-1}{x} \cdot (1-p)^x p^r = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)}(1-p)^x p^r$$

where $r > 0$ and $p \in (0, 1)$ are parameters.

The Normal distribution

If the real x has a Normal distribution with parameters μ and σ^2, its density is given by

$$\pi(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

We write $x \mid \mu, \sigma^2 \sim \text{Normal}(\mu, \sigma^2)$ and $\pi(x \mid \mu, \sigma^2) = \text{Normal}(x; \mu, \sigma^2)$.

The Poisson distribution

If $x \in \{0, 1, 2, \dots\}$ has Poisson distribution with parameter $\lambda > 0$ then the probability mass function is

$$e^{-\lambda}\frac{\lambda^x}{x!}.$$

We write $x \mid \lambda \sim \text{Poisson}(\lambda)$ and $\pi(x \mid \lambda) = \text{Poisson}(x; \lambda)$. The Poisson distribution has expectation λ and variance λ.

Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MVE550 Stochastic Processes and Bayesian Inference
Exam January 8 2024

1. (a) We get, using the information in the appendix about the Geometric distribution,

$$\mu = \mathrm{E}[X] = \mathrm{E}[Y-1] = \mathrm{E}[Y] - 1 = \frac{1}{p} - 1$$

and

$$\sigma^2 = \mathrm{Var}[X] = \mathrm{Var}[Y-1] = \mathrm{Var}[Y] = \frac{1-p}{p^2}.$$

(b) The process is critical when $1/p - 1 = 1$, i.e., when $p = 1/2$. When $p > 1/2$ the process is sub-critical and when $p < 1/2$ it is supercritical. To compute in that last case, we start with finding the probability generating function:

$$\begin{aligned} G_x(s) &= \mathrm{E}\left[s^X\right] = \mathrm{E}\left[s^{Y-1}\right] \\ &= \sum_{y=1}^{\infty} p(1-p)^{y-1} s^{y-1} \\ &= p\sum_{x=0}^{\infty} ((1-p)s)^x \\ &= \frac{p}{1-(1-p)s}. \end{aligned}$$

Finding the smallest positive root of the equation $G_X(s) = s$ gives (using that the equation must have $s = 1$ as a root):

$$\begin{aligned} \frac{p}{1-(1-p)s} &= s \\ p &= s - s^2(1-p) \\ s^2(1-p) - s + p &= 0 \\ (s-1)(s(1-p)-p) &= 0 \end{aligned}$$

so the smallest positive root is found setting $s(1-p) - p = 0$. Summing up, we get that the probability of extinction is 1 if $p \geq 1/2$, and if $p < 1/2$ it is $p/(1-p)$.

(c) Using the law of total variance, writing $X_1, \ldots, X_n$ for independent copies of X, and doing computations stepwise, we get

$$\begin{aligned}
\operatorname{Var}[Z_2] &= \operatorname{Var}[\operatorname{E}[Z_2 \mid Z_1]] + \operatorname{E}[\operatorname{Var}[Z_2 \mid Z_1]] \\
&= \operatorname{Var}\left[\operatorname{E}\left[\sum_{i=1}^{n} X_i \mid Z_1 = n\right]\right] + \operatorname{E}\left[\operatorname{Var}\left[\sum_{i=1}^{n} X_i \mid Z_1 = n\right]\right] \\
&= \operatorname{Var}[n \operatorname{E}[X_i] \mid Z_1 = n] + \operatorname{E}[n \operatorname{Var}[X_i] \mid Z_1 = n] \\
&= \operatorname{Var}[Z_1 \operatorname{E}[X]] + \operatorname{E}[Z_1 \operatorname{Var}[X]] \\
&= \operatorname{Var}[Z_1 \mu] + \operatorname{E}\left[Z_1 \sigma^2\right] \\
&= \mu^2 \operatorname{Var}[Z_1] + \sigma^2 \operatorname{E}[Z_1] \\
&= \mu^2 \operatorname{Var}[X] + \sigma^2 \operatorname{E}[X] \\
&= \mu^2 \sigma^2 + \mu \sigma^2 \\
&= (\mu + 1)\mu\sigma^2 \\
&= (1/p - 1 + 1)(1/p - 1)\frac{1 - p}{p^2} \\
&= \frac{(1 - p)^2}{p^4}
\end{aligned}$$

2. (a) The lines P_1, P_2, P_3, and P_4 could be modelled with independent Dirichlet distributions as follows:

$$\begin{aligned}
P_1 &\sim \text{Dirichlet}(1, 1, 1, 1) \\
P_2 &\sim \text{Dirichlet}(1, 1, 1, 1) \\
P_3 &\sim \text{Dirichlet}(0, 1, 1, 1) \\
P_4 &\sim \text{Dirichlet}(0, 0, 1, 1)
\end{aligned}$$

(other distribution parameters than the given ones could be used, except for the zero values, which must be zero to encode the impossibility of certain jumps).

(b) There is only one recorded transition out of state 4, that is, a transition to state 3. Thus, including the pseudo counts 0,0,1,1 from the prior, we get the posterior

$$P_4 \mid \text{data} \sim \text{Dirichlet}(0, 0, 2, 1).$$

As $X_{11} \mid P_4, X_{10}$ is Multinomial (in fact Binomial as it can have only two possible values), we get, as in the course examples, the predictive distribution

$$\begin{aligned}
\Pr[X_{11} = 3 \mid \text{data}] &= \frac{2}{3} \\
\Pr[X_{11} = 4 \mid \text{data}] &= \frac{1}{3}
\end{aligned}$$

(c) In the original exam, the indexation started at $i = 0$, while the illustration and data table assumed a start at $i = 1$. The solution below assumes that the indexation starts at $i = 1$; answering with formulas starting at $i = 0$ was of course also acceptable. Note that factors in $\pi(\text{data} \mid \lambda)$ that concern how X_i depends on X_{i-1} do not contain λ, and they can thus be dropped below.

$$\begin{aligned}
\pi(\lambda \mid \text{data}) &\propto_\lambda \pi(\text{data} \mid \lambda)\pi(\lambda) \\
&\propto_\lambda \prod_{i=1}^{n} \text{Exponential}(y_i; \lambda x_i) \cdot \frac{1}{\lambda} \\
&\propto_\lambda \frac{1}{\lambda} \prod_{i=1}^{n} \lambda x_i \exp(-\lambda x_i y_i) \\
&\propto_\lambda \frac{1}{\lambda} \lambda^n \exp\left(-\sum_{i=1}^{n} \lambda x_i y_i\right) \\
&\propto_\lambda \lambda^{n-1} \exp\left(-\lambda \sum_{i=1}^{n} x_i y_i\right) \\
&\propto_\lambda \text{Gamma}\left(\lambda; n, \sum_{i=1}^{n} x_i y_i\right),
\end{aligned}$$

so that the posterior distribution for λ must be Gamma $\left(n, \sum_{i=1}^{n} x_i y_i\right)$.

3. (a) Let us start for example with T_{12}. To go to state 2 from state 1, the MH algorithm needs to first propose state 2, and then accept this proposal. The probability of proposing 2 is $1/2$, by looking at the undirected weighted graph, while the acceptance probability, for any transition from i to j is

$$\min\left(1, \frac{\pi(j)}{\pi(i)}\right)$$

as the proposals are symmetric. Thus we get

$$T_{12} = \frac{1}{2} \cdot \frac{0.1}{0.8} = \frac{1}{16}$$

We get the similar result

$$T_{13} = \frac{1}{2} \cdot \frac{0.1}{0.8} = \frac{1}{16}$$

and thus

$$T_{11} = 1 - T_{12} - T_{13} = \frac{7}{8}$$

We then also get

$$
\begin{aligned}
T_{21} &= \frac{1}{2} \cdot 1 = \frac{1}{2} \\
T_{23} &= \frac{1}{2} \cdot 1 = \frac{1}{2} \\
T_{22} &= 1 - T_{21} - T_{23} = 0 \\
T_{31} &= \frac{1}{2} \cdot 1 = \frac{1}{2} \\
T_{32} &= \frac{1}{2} \cdot 1 = \frac{1}{2} \\
T_{33} &= 1 - T_{31} - T_{32} = 0
\end{aligned}
$$

so that in the end

$$T = \begin{bmatrix} \frac{7}{8} & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

(b) Any Markov chain produced by a Metropolis Hastings algorithm is time reversible.

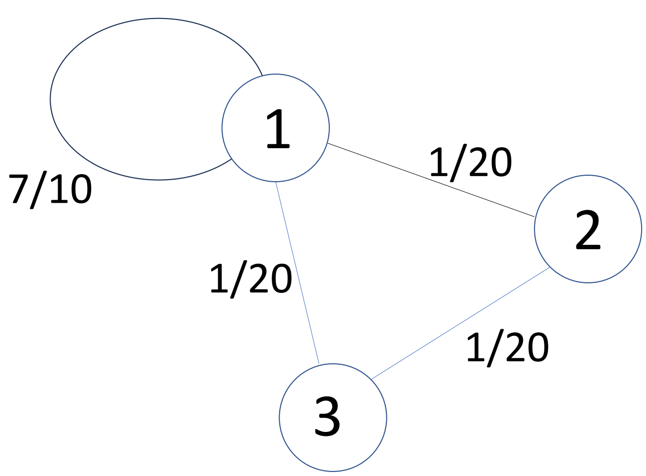

Figure 1: The answer to question 3c.

(c) As the stationary distribution is $q = (0.8, 0.1, 0.1)$ we get

$$\begin{aligned}
w_{12} &= q_1 T_{12} = 0.8\frac{1}{16} = \frac{1}{20} = w_{21} \\
w_{13} &= q_1 T_{13} = 0.8\frac{1}{16} = \frac{1}{20} = w_{31} \\
w_{23} &= q_2 T_{23} = 0.1\frac{1}{2} = \frac{1}{20} = w_{32} \\
w_{11} &= q_1 T_{11} = 0.8\frac{7}{8} = \frac{7}{10} \\
w_{22} &= q_2 T_{22} = 0 \\
w_{33} &= w_3 T_{33} = 0
\end{aligned}$$

and the weighted undirected graph of Figure 1. Note that multiplying all weights with the same constant, e.g., 20, gives a solution that works equally well.

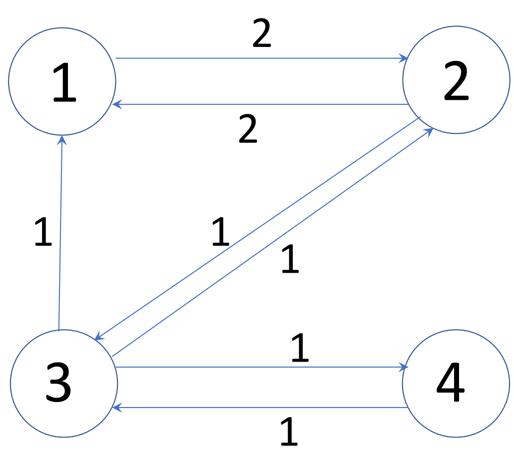

Figure 2: The answer to question 4a.

4. (a) See Figure 2.

(b) We get

$$Q = \begin{bmatrix} -2 & 2 & 0 & 0 \\ 2 & -3 & 1 & 0 \\ 1 & 1 & -3 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

and

$$\tilde{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

(c) The chain is irreducible as can be seen from the transition rate graph drawn in (a). Thus it is ergodic, and has a unique limiting distribution which can be found as the unique stationary distribution satisfying $vQ = 0$. To find this stationary distribution we replace the first column of Q with 1's to obtain Q', and then solve the equation $vQ' = (1, 0, 0, 0)$. In other words, we can compute

$$v = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ 1 & 1 & -3 & 1 \\ 1 & 0 & 1 & -1 \end{bmatrix}^{-1} .$$

In fact, it is not too difficult to show that

$$v = \left(\frac{5}{13}, \frac{4}{13}, \frac{2}{13}, \frac{2}{13} \right).$$

(d) We can make both states 3 and 4 into a single absorbing state (or we can notice that state 4 is reachable from state 2 only via state 3, so we can ignore state 4 in this question). From Q we get the fundamental matrix

$$F = -\begin{bmatrix} -2 & 2 \\ 2 & -3 \end{bmatrix}^{-1} .$$

The answer to the question is the sum of the second row of F, which can be computed as

$$-\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 2 & -3 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

In fact, the answer becomes 2.

5. (a) The time before the first move is Exponentially distributed with parameter $\lambda = 0.3$. The probabilty that such a variable is larger than 2 can be computed in R with

```
1- pexp(2, 0.3)
```

Mathematically, the cumulative distribution function is $F(x) = 1 - \exp(-\lambda x)$, so we may also compute the result as

$$\exp(-2 \cdot 0.3) = 0.548816.$$

(b) The 14 movements are each placed uniformly in the time interval [0,20]. The probability that there are no movements in the interval $[0, 2]$ can then be computed as the probability that all movements are in the interval $[2, 20]$:

$$\left(\frac{20 - 2}{20} \right)^{14} = 0.2287679.$$

(c) The count of single-cog turns forward can be viewed as a Poisson process with parameter $\lambda = 0.3$. The arrival time for the n'th arrival is then distributed as Gamma(n, λ). In other words, the probability that arrival number $3 \cdot 14 = 42$ occurs before time 100 can be computed with

```
pgamma(100, 14*3, 0.3)
```

in R. A mathematical expression for the same thing is

$$\int_0^{100} \mathrm{Gamma}(t, 42, 0.3)\, dt = \int_0^{100} \frac{0.3^{42}}{\Gamma(42)} t^{41} \exp(-0.3t)\, dt.$$

The numerical answer is 0.02210704. Another way to compute this is with

```
1 - ppois(41, 300*0.3)
```

or the corresponding mathematical expression.