DATA 1010
In-class exercises
Samuel S. Watson
19 October 2018

Problem 1

The **geometric distribution** with parameter $p \in (0,1]$ is the distribution of the index of the first success in a sequence of independent Bernoulli trials.

Find the probability mass function of the geometric distribution.

Solution

The probability that the first success occurs on trial k is equal to the probability that the first $k-1$ trials fail and the kth trial succeeds. The probability of this event is $p(1-p)^{k-1}$. Therefore, the probability mass function of the geometric distribution is

$$m(k) = p(1-p)^{k-1}.$$

Problem 2

Use Monte Carlo to find the mean and variance of the geometric distribution with parameter $p = 1/3$.

Solution

Here's an example solution:

```
using Statistics

function sample_geometric(p)
    k = 1
    while true
        if rand() < p
            return k
        else
            k += 1
        end
    end
end

samples = [sample_geometric(1/3) for i=1:1_000_000]

m = mean(samples)
σ² = mean(x^2 for x in samples) - m^2

(m, σ²)
```

Problem 3

(i) Find the expected value of S, where S is a sum of 1000 independent Bernoulli random variables with success probability $p = \frac{3}{1000}$.

(ii) Find the probability mass function of S. Hint: find an expression representing the probability mass at each k from 0 to 1000, and then use Julia to evaluate it. You will need to define `n = big(1000)` and `p = big(3)/1000` because arbitrary precision arithmetic is required to avoid overflow issues.

(iii) Compare your results to the probability mass function $m(k) = \frac{3^k}{k!}e^{-3}$ defined on $\{0, 1, 2, \ldots\}$.

Solution

(ii) Consider all 2^{1000} possible length-1000 strings of 0's or 1's. Of these, there are $\binom{1000}{k}$ with k ones and $1000-k$ zeros, and each of those $\binom{1000}{k}$ strings has a probability of $p^k(1-p)^{1000-k}$ of being the result of independent sequence of Bernoulli(p) random variables (where $p = \frac{3}{1000}$). Therefore, the probability of the event $\{S = k\}$ is $\binom{1000}{k}p^k(1-p)^{1000-k}$. We can obtain a vector of these probabilities as follows:

```
n = big(1000)
p = big(3)/1000
massfunction = [binomial(n,k)*p^k*(1-p)^(n-k) for k=0:1000]
```

(iii) We can run `[3^big(k)/factorial(big(k))*exp(-3) for k=0:1000]` to get the first 1001 values of the given probability mass function. We see that the values are quite similar. The first ten pairs of values are

```
(0.0495631, 0.0497871)
(0.149137, 0.149361)
(0.224154, 0.224042)
(0.224379, 0.224042)
(0.168284, 0.168031)
(0.100869, 0.100819)
(0.0503334, 0.0504094)
(0.0215065, 0.021604)
(0.00803259, 0.00810151)
(0.0026641, 0.0027005)
```

We can see that they indeed match quite closely.

Problem 4

Suppose $\lambda > 0$, and find the mean and variance of a sum of n independent Bernoulli random variables with parameter $p = \lambda/n$ (where $n > \lambda$). Use your results to posit values for the expectation and variance of a Poisson random variable with parameter λ.

Solution

The average number of successes in n Bernoulli(λ/n) trials is $(n)(\lambda/n) = \lambda$, by linearity of expectation. Therefore, we expect that the mean of a Poisson random variable with parameter λ is equal to λ.

Similarly, the variance of the number of successes in n Bernoulli(λ/n) trials is equal to $n\frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right) = \lambda(1-\lambda/n)$. Taking $n \to \infty$, we predict that the variance of a Poisson random variable with parameter λ is also equal to λ.

Problem 5

Suppose that the number of typos on a page is a Poisson random variable with mean $\lambda = \frac{1}{3}$.

(i) Provide an explanation for why the Poisson distribution might be a good approximation for the distribution of typos on a page.

(ii) Find the probability that a particular page is typo-free.

Solution

(i) A typo opportunities on a page convert to actual typos with a small but roughly constant probability, there are quite a few of them, and different typos are (roughly) independent of one another. Thus the number of typos is a sum of independent Bernoulli random variables.

(ii) The probability that a Poisson random variable with parameter $\lambda = \frac{1}{3}$ is equal to 0 is

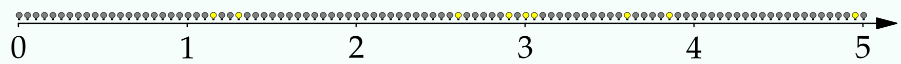
$$\frac{\lambda^0}{0!}e^{-\lambda} = e^{-1/3} \approx 71.6\%.$$

Problem 6

Imagine placing a light bulbs activated by independent Bernoulli(λ/n) random variables at every multiple of $1/n$ on the positive real number line. Consider the position X_n of the **leftmost lit bulb**.

(i) For each $t > 0$, find the limit as $n \to \infty$ of $\mathbb{P}(X_n > t)$.

(ii) Find the PDF associated with the measure that you found in part (a).



Solution

(i) The probability that it occurs to the right of a point $t > 0$ is equal to the probability that all of the $\lfloor nt \rfloor$ bulbs to the left remain unlit:

$$\mathbb{P}(X > t) = \left(1 - \frac{\lambda}{n}\right)^{nt}$$

This probability converges to $\mathrm{e}^{-\lambda t}$ as $n \to \infty$.

(ii) To find the PDF in terms of the CDF, we note that the general relationship between a PDF f and its CDF F is

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

By the fundamental theorem of calculus, this implies that $F'(x) = f(x)$. Thus the density associated with the measure

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}[1 - \mathrm{e}^{-\lambda x}] = \lambda \mathrm{e}^{-\lambda x}.$$