4. The Poisson Distribution

The Probability Density Function

We have shown that the kth arrival time in the Poisson process has the gamma probability density function with shape parameter k and rate parameter r:

$$f^{*k}(t) = r^k \frac{t^{k-1}}{(k-1)!} e^{-r\,t}, \quad t \geq 0$$

Recall also that at least k arrivals come in the interval $(0, t]$ if and only if the kth arrival occurs by time t:

$$(N_t \geq k) \Leftrightarrow (T_k \leq t)$$

1. Use integration by parts to show that

$$\mathbb{P}(N_t \geq k) = \int_0^t f^{*k}(s)\,ds = 1 - \sum_{j=0}^{k-1} e^{-r\,t} \frac{(r\,t)^j}{j!}, \quad k \in \mathbb{N}$$

2. Use the result of Exercise 1 to show that the probability density function of the number of arrivals in the interval $(0, t]$ is

$$\mathbb{P}(N_t = k) = e^{-r\,t} \frac{(r\,t)^k}{k!}, \quad k \in \mathbb{N}$$

The corresponding distribution is called the **Poisson distribution** with parameter $r\,t$; the distribution is named after Simeon Poisson.

3. In the Poisson experiment, vary r and t with the scroll bars and note the shape of the density function. Now with $r = 2$ and $t = 3$, run the experiment 1000 times with an update frequency of 10 and watch the apparent convergence of the relative frequency function to the density function.

The Poisson distribution is one of the most important in probability. In general, a discrete random variable N in an experiment is said to have the Poisson distribution with parameter $c > 0$ if it has the probability density function

$$g(k) = e^{-c} \frac{c^k}{k!}, \quad k \in \mathbb{N}$$

4. Show directly that g is a valid probability density function.

5. Show that

   a. $g(n-1) < g(n)$ if and only if $n < c$.
   b. g at first increases and then decreases, and thus the distribution is unimodal
   c. If c is not an integer, there is a single mode at $\lfloor c \rfloor$. If c is an integer there are two modes at $c - 1$ and c.

6. Suppose that requests to a web server follow the Poisson model with rate $r = 5$. per minute. Find the probability that there will be at least 8 requests in a 2 minute period.

7. Defects in a certain type of wire follow the Poisson model with rate 1.5 per meter. Find the probability that there will be no more than 4 defects in a 2 meter piece of the wire.

Moments

Suppose that N has the Poisson distribution with parameter c. The following exercises give the mean, variance, and probability generating function of N.

8. Show that $\mathbb{E}(N) = c$.

9. Show that $\text{var}(N) = c$.

10. Show that $\mathbb{E}(u^N) = e^{c(u-1)}$. for $u \in \mathbb{R}$.

Returning to the Poisson process $\{N_t : t \geq 0\}$ with rate parameter r, it follows that $\mathbb{E}(N_t) = r\,t$ and $\text{var}(N_t) = r\,t$ for $t \geq 0$. Once again, we see that r can be interpreted as the average arrival *rate*. In an interval of length t, we expect about $r\,t$ arrivals.

11. In the Poisson experiment, vary r and t with the scroll bars and note the location and size of the mean/standard deviation bar. Now with $r = 3$ and $t = 4$, run the experiment 1000 times with an update frequency of 10 and watch the apparent convergence of the sample mean and standard deviation to the distribution mean and standard deviation, respectively.

12. Suppose that customers arrive at a service station according to the Poisson model, at a rate of $r = 4$. Find the mean and standard deviation of the number of customers in an 8 hour period.

Stationary, Independent Increments

Let us see what the basic regenerative assumption of the Poisson process means in terms of the counting variables $\{N_t : t \geq 0\}$.

13. Show that if $s < t$, then $N_t - N_s$ is the number of arrivals in the interval $(s, t]$.

Recall that our basic assumption is that the process essentially starts over at time s and the behavior after time s is independent of the behavior before time s.

14. Argue that:

   a. $N_t - N_s$ has the same distribution as N_{t-s} namely Poisson with parameter $r\,(t - s)$.
   b. $N_t - N_s$ and N_s are independent.

15. Suppose that N and M are independent random variables, and that N has the Poisson distribution with parameter c and M has the Poisson distribution with parameter d. Show that $N + M$ has the Poisson distribution with parameter $c + d$.

   a. Give a probabilistic proof, based on the Poisson process.
   b. Give an analytic proof using probability density functions.
   c. Give an analytic proof using probability generating functions.

16. In the Poisson experiment, select $r = 1$ and $t = 3$. Run the experiment 1000 times, updating after each run. By computing the appropriate relative frequency functions, investigate empirically the independence of the random variables N_1 and $N_3 - N_1$.

Normal Approximation

Now note that for $k \in \mathbb{N}_+$,

$$N_k = N_1 + (N_2 - N_1) + \cdots + (N_k - N_{k-1})$$

The random variables in the sum on the right are independent and each has the Poisson distribution with parameter r.

17. Use the central limit theorem to show that the distribution of the standardized variable below converges to the standard normal distribution as $k \to \infty$.

$$Z_k = \frac{N_k - k\,r}{\sqrt{k\,r}}$$

A bit more generally, the same result is true with the integer k replaced by the positive real number c. Thus, if N has the Poisson distribution with parameter c, and c is "large", then the distribution of N is approximately normal with mean c and standard deviation $\sqrt{c}$. When using the normal approximation, we should remember to use the continuity correction, since the Poisson is a discrete distribution.

18. In the Poisson experiment, set $r = 1$ and $t = 1$. Increase r and t and note how the graph of the probability density function becomes more bell-shaped.

19. In the Poisson experiment, set $r = 5$ and $t = 4$. Run the experiment 1000 times with an update frequency of 100. Compute and compare the following:

   a. $\mathbb{P}(15 \leq N_4 \leq 22)$
   b. The relative frequency of the event $\{15 \leq N_4 \leq 22\}$.
   c. The normal approximation to $\mathbb{P}(15 \leq N_4 \leq 22)$.

20. Suppose that requests to a web server follow the Poisson model with rate $r = 5$. Compute the normal approximation to the probability that there will be at least 280 requests in a 1 hour period.

Conditional Distributions

Consider again the Poisson model with arrival time sequence $(T_1, T_2, ...)$ and counting process $\{N_t : t \geq 0\}$.

⊡ 21. Let $t > 0$. Show that the conditional distribution of T_1 given $N_t = 1$ is uniform on the interval $(0, t)$. Interpret the result.

⊡ 22. More generally, show that given $N_t = n$, the conditional distribution of $(T_1, ..., T_n)$ is the same as the distribution of the order statistics of a random sample of size n from the uniform distribution on the interval $(0, t)$.

Note that the conditional distribution in the last exercise is independent of the rate r. This result means that, in a sense, the Poisson model gives the most "random" distribution of points in time.

⊡ 23. Suppose that requests to a web server follow the Poisson model, and that 1 request comes in a five minute period. Find the probability that the request came during the first 3 minutes of the period.

⊡ 24. In the Poisson experiment, set $r = 1$ and $t = 2$. Run the experiment 1000 times, updating after each run. Compute the appropriate relative frequency functions and investigate empirically the theoretical result in Exercise 23.

⊡ 25. Suppose that $0 < s < t$ and that n is a positive integer. Show that the conditional distribution of N_s given $N_t = n$ is binomial with trial parameter n and success parameter $p = \frac{s}{t}$. Note that the conditional distribution is independent of the rate r. Interpret the result.

⊡ 26. Suppose that requests to a web server follow the Poisson model, and that 10 requests come during a 5 minute period. Find the probability that at least 4 requests came during the first 3 minutes of the period.

Estimating the Rate

In many practical situations, the rate r of the process in unknown and must be estimated based on observing the number of arrivals in an interval.

⊡ 27. Show that $\mathbb{E}\left(\frac{N_t}{t}\right) = r$ and hence $\frac{N_t}{t}$ is an unbiased estimator of r.

Since the estimator is unbiased, the variance measures the mean square error of the estimator.

⊡ 28. Show that $\text{var}\left(\frac{N_t}{t}\right) = \frac{r}{t}$ and hence $\text{var}\left(\frac{N_t}{t}\right) \to 0$ as $t \to \infty$. This means that $\frac{N_t}{t}$ is an consistent estimator of r.

⊡ 29. In the Poisson experiment, set $r = 3$ and $t = 5$. Run the experiment 100 times, updating after each run.

    a. For each run, compute the estimate of r based on N_t.
    b. Over the 100 runs, compute the average of the squares of the errors.
    c. Compare the result in (b) with the variance in Exercise 28.

⊡ 30. Suppose that requests to a web server follow the Poisson model with unknown rate r per minute. In a one hour

period, the server receives 342 requests. Estimate r.