ISYE 6420: Homework 2

aelhabr3

1. Cell Clusters in 3D Petri Dishes

Instructions

The number of cell clusters in a 3D Petri dish has a Poisson distribution with mean $\lambda = 5$. *Find the percentage of Petri dishes that have (a) 0 clusters, (b) at least one cluster, (c) more than 8 clusters, and (d) between 4 and 6 clusters inclusive.*

Response

Let X represent the number of cell clusters. The instructions indicate that $X \sim \text{Pois}(\lambda = 5)$. Thus the probability mass function (PMF) $f(x)$ of X is as follows.

$$f(x) = \frac{\lambda x e^{-\lambda}}{x!} = \frac{5^x e^{-5}}{x!}.$$

Note that $f(x) = P(X = x)$.

The cumulative distribution function (CDF) $F(x)$ of X is as follows.

$$F(x) = e^{-\lambda} \sum_{k=0}^{x} \frac{\lambda^k}{k!} = e^{-5} \sum_{k=0}^{x} \frac{5^k}{k!}.$$

Note that $F(x) = P(X \leq x)$.

Now we address each part of the instructions.

(a)

We use the PDF to find the percentage of Petri dishes having 0 clusters.

$$P(X = 0) = \frac{5^0 e^{-5}}{0!} = e^{-5} = 0.0067$$

We can check this number using `R` as follows.

```
l <- 5
p_q1a <- ppois(0, lambda = l, lower.tail = T)
p_q1a
```

```
## [1] 0.006737947
```

Thus, we find that the percentage is 100 * 0.0067 = 0.6738%.

(b)

Next, we deduce the percentage of Petri dishes having at least one cluster from our finding in (a).

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (e^{-5}) = 0.9933$$

Again, we can check our number with `R`.

```
1 - p_q1a
```

```
## [1] 0.9932621
```

```
# Or
p_q1b <- ppois(0, lambda = 1, lower.tail = F)
p_q1b
```

```
## [1] 0.9932621
```

Thus, we find that the percentage is 99.3262%.

(c)

To find the percentage of Petri dishes having more than 8 clusters, we use the CDF.

$$P(X > 8) = 1 - P(X \leq 8) = 1 - (e^{-5} \sum_{k=0}^{8} \frac{5^k}{k!}).$$

Let's use `R` to calculate this (since writing out all of the individual terms in the sum would be tedious).

```
p_q1c <- ppois(8, lambda = 1, lower.tail = F)
p_q1c
```

```
## [1] 0.06809363
```

We find that the percentage is 6.8094%.

(d)

Finally, to find the percentage of Petri dishes having between 4 and 6 clusters inclusive, we use the difference in CDFs. (Note that we must be careful with the lower bound 4, which we want to include.)

$$P(4 \leq X \leq 6) = P(X \leq 6) - P(X < 3) = (e^{-5}\sum_{k=0}^{6}\frac{5^k}{k!}) - (e^{-5}\sum_{k=0}^{3}\frac{5^k}{k!}) = e^{-5}\sum_{k=4}^{6}\frac{5^k}{k!}.$$

```
p_q1d <- ppois(6, lambda = 1, lower.tail = T) - ppois(4 - 1, lambda = 1, lower.tail = T)
p_q1d
```

```
## [1] 0.4971575
```

We find that the percentage is 49.7158%.

2. Silver-Coated Nylon Fiber

Instructions

Silver-coated nylon fiber is used in hospitals for its anti-static electricity properties, as well as for antibacterial and antimycotic effects. In the production of silver-coated nylon fibers, the extrusion process is interrupted from time to time by blockages occurring in the extrusion dyes. The time in hours between blockages, T, has an exponential $\mathcal{E}$(1/10) distribution, where 1/10 is the rate parameter.
Find the probabilities that
(a) a run continues for at least 10 hours,
(b) a run lasts less than 15 hours, and
(c) a run continues for at least 20 hours, given that it has lasted 10 hours.

If you use software, be careful about the parametrization of exponentials.

Response

The instructions indicate that $T \sim \mathcal{E}(1/\lambda = 1/10)$. (Note that λ represents a rate. It is like the inverse of λ for the Poisson distribution.) The probability density function (PDF) and CDF are as follows. (Note that the PDF is appropriate for continuous distributions, while the PMF is appropriate for discrete distributions.)

$$f(x) = \lambda e^{-\lambda x} = -\frac{1}{10}e^{-x/10}, x \geq 0.$$

$$F(x) = 1 - e^{-\lambda x} = 1 - e^{-x/10}, x \geq 0.$$

(a)

We find the probability of a run continuing for at least 10 hours as follows.

$$P(X > 10) = 1 - P(X \leq 10) = 1 - (1 - e^{-(10)/(10)}) = e^{-1} = 0.3679.$$

We can check this with `R`.

```
exp(-1)
```

```
## [1] 0.3678794
```

```
# Or
l <- 10
r <- 1 / l
p_q2a <- pexp(10, rate = r, lower.tail = F)
p_q2a
```

```
## [1] 0.3678794
```

The probability is 0.3679.

(b)

We find the probability that a run lasts less than 15 hours as follows.

```
1 - pexp(15, rate = r, lower.tail = F)
```

```
## [1] 0.7768698
```

```
# Or
p_q2b <- pexp(15, rate = r, lower.tail = T)
p_q2b
```

```
## [1] 0.7768698
```

The probability is 0.7769.

(c)

Note that the $\mathcal{E}$ distribution has the memoryless property, which means that $P(X > (s+t)|X > s) = P(X > t)$ for $s, t > 0$.

Now we calculate the probability that run continues for at least 20 hours, given that it has lasted 10 hours, as follows.

$$P(X > 20|X > 10) = P(X > 10) = e^{-10/10} = e^{-1} = 0.3679.$$

Checking with `R`, we find the following.

```
exp(-1)
```

```
## [1] 0.3678794
```

```
# Or
p_q2c <- pexp(20 - 10, rate = r, lower.tail = F)
p_q2c
```

```
## [1] 0.3678794
```

The probability is 0.3679.

3. 2-D Density Tasks

Instructions

If

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y} & 0 \le x \le y, \quad \lambda > 0 \\ 0, & \text{else.} \end{cases}$$

Show that:
(a) marginal distribution $f_X(x)$ *is exponential* $\mathcal{E}(\lambda)$*.*
(b) marginal distribution $f_Y(y)$ *is Gamma* $\mathcal{G}a(2,\lambda)$*.*
(c) conditional distribution $f(y|x)$ *is shifted exponential* $f(y|x) = \lambda e^{-\lambda(y-x)}$*,* $y \ge x$*.*
(d) conditional distribution $f(x|y)$ *is uniform* $\mathcal{U}(0,y)$*.*

Response

(a)

$$
\begin{aligned}
f_X(x) &= \int_x^\infty \lambda^2 e^{-\lambda y} dy \\
&= \lambda^2 \int_x^\infty e^{-\lambda y} dy \\
&= \lambda^2 (\frac{1}{-\lambda} e^{-\lambda y}\Big|_x^\infty) \\
&= -\lambda(0 - e^{-\lambda x}) \\
&= \lambda e^{-\lambda x}
\end{aligned}
$$

Note that the PDF of $\mathcal{E}(\lambda)$ is $f(x) = \lambda e^{-\lambda x}$, which matches the result above.

(b)

$$
\begin{aligned}
f_Y(y) &= \int_{x0}^y \lambda^2 e^{-\lambda y} dx \\
&= \lambda^2 e^{-\lambda y} \int_0^y dx \\
&= \lambda^2 e^{-\lambda y} x\Big|_0^y \\
&= \lambda^2 e^{-\lambda y} y
\end{aligned}
$$

Note that general form of the PDF of $\mathcal{G}a(\alpha, \beta)$ is as follows.

$$
f(x; \alpha, \beta) = P(x = X) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad x, \alpha, \beta > 0.
$$

Then, if we substitute λ for β and 2 for α, and we evaluate the denominator as $\Gamma(2) = (2-1)! = 1$ (given the fact that $\Gamma(n) = (n-1)! \quad \forall \mathbb{Z}^+$), then we see how the formulation above matches the distribution that we should see–$\mathcal{G}a(2, \lambda)$.

$$
\begin{aligned}
f(y) &= \frac{\lambda^2 y^{2-1} e^{-\lambda y}}{\Gamma(2)} \\
&= \frac{\lambda^2 y e^{-\lambda y}}{1} \\
&= \lambda^2 y e^{-\lambda y} \sim \mathcal{G}a(2, \lambda)
\end{aligned}
$$

(c)

Showing that $f(y|x)$ is the shifted exponential $f(y|x) = \lambda e^{-\lambda(y-x)}$ for $y \geq x$ follows directly from the result of (a) and the given PDF $f_{X,Y}(x, y)$ for $y \geq x$.

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\
&= \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} \\
&= \lambda e^{-\lambda(y+(-x))} \\
&= \lambda e^{-\lambda(y-x)}.
\end{aligned}
$$

(Note that we are being "verbose" by writing X and Y in the subscript of terms that involve both, e.g. $f_{X,Y}(x,y)$ instead of just $f(x,y)$.)

(d)

We solve for $f_{X|Y}(x|y)$ using the result from (b) and the given PDF as follows.

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\
&= \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} \\
&= \frac{1}{y}.
\end{aligned}
$$

Note that the general form of the PDF of $\mathcal{U}(a,b)$ is $\frac{1}{b-a}$ (for $x \in [a,b]$), which matches the result above (if we substitute $b = y$ and $a = 0$).

4. Nylon Fiber Continued

Instructions

In the Exercise 2, the times (in hours) between blockages of the extrusion process, T, had an exponential $\mathcal{E}(\lambda)$ distribution. Suppose that the rate parameter λ is unknown, but there are three measurements of interblockage times, $T_1 = 3, T_2 = 13$, and $T_3 = 8$.

(a) How would a classical statistician estimate λ?
(b) What is the Bayes estimator of λ if the prior is $\pi(\lambda) = \frac{1}{\sqrt{\lambda}}$, $\lambda > 0$?

Hint: In (b) the prior is not a proper distribution, but the posterior is. Identify the posterior from the product of the likelihood from (a) and the prior, no need to integrate.

Response

(a)

A classical statistician would estimate λ using the MLE, which is $\hat{\lambda}_{MLE} = \frac{1}{\bar{X}}$.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} T_i = \frac{1}{3}(3+13+8) = 8.$$

Thus,

$$\hat{\lambda}_{MLE} = \frac{1}{\overline{X}} = \frac{1}{8}.$$

(b)

The instructions indicate that we are trying to estimate the rate $T_1, \ldots, T_n \sim \mathcal{E}(\theta)$ given the prior $\pi(\lambda) = \theta = \frac{1}{\sqrt{\lambda}} = \lambda^{-0.5}, \lambda > 0$.

Thus, we have the following problem setup.

$$\begin{array}{rcccc} \text{Likelihood} & : & f(t|\theta) & \sim & \theta^n e^{-\theta \sum_{i=1}^{n} x_i} \\ \text{Prior} & : & \pi(\lambda) = \theta & \sim & \lambda^{-1/2} \end{array}$$

Then,

$$\begin{array}{rcccc} \text{Posterior} & \propto & \text{Likelihood} & \times & \text{Prior} \\ \pi(\theta|t) & \propto & f(t|\theta) & \times & \theta \\ & \propto & \theta^n e^{-\theta \sum_{i=1}^{n=3} t_i} & \times & \lambda^{-0.5} \\ & \propto & \lambda^{n-0.5} \theta^n e^{-\theta \sum_{i=1}^{n} t_i} & & \\ & \propto & \lambda^{(n+0.5)-1} \theta^n e^{-\theta \sum_{i=1}^{n} t_i}. & & \end{array}$$

Thus, we see that the posterior follows a $\mathcal{G}a(\alpha, \beta)$ distribution with $\alpha = n + 0.5, \beta = \sum_{i=1}^{n} t_i$. Given that $n = 3$ and $\sum_{i=1}^{n} t_i = 24$, we can say that posterior is $\mathcal{G}a(3.5, 24)$.

We can go further with this result and calculate the posterior mean.

$$\mathrm{E}[T|\alpha, \beta] \quad = \quad \tfrac{\alpha}{\beta} = \tfrac{3.5}{24} = \tfrac{7}{48} \approx 0.1458.$$

We note that the posterior mean is relatively close to the MLE estimate $\frac{1}{8} = 0.125$.