Worked examples on distributions of continuous random variables

Lily Yen

Due February 16th, 2015

Example 1 The time spent by a randomly selected internet user who uses an internet service provider has a gamma distribution with mean 20 min and variance 80 min^2.

1. What are the values of the parameters α and β?
2. What is the probability that a user spends at most 25 min connected to the internet?
3. What percentage of users spend between 15 and 30 min connected to the internet?

Solution 1.

1. For the gamma distribution, $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$. Therefore $\frac{\sigma^2}{\mu} = \frac{\alpha\beta^2}{\alpha\beta} = \beta$. In the present case, $\beta = \frac{80\,\text{min}^2}{20\,\text{min}} = 4.0\,\text{min}$ and $\alpha = \frac{\mu}{\beta} = \frac{20\,\text{min}}{4.0\,\text{min}} = 5.0$.

2. The pdf for the gamma distribution is $f(x;\alpha,\beta) = \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}$ $(x > 0)$. If the present case

$$f(x;5,4) = \frac{1}{4^5(5-1)!}x^4e^{-x/4} = \frac{1}{24576}x^4e^{-x/4}.$$

The desired probability is then

$$\int_0^{25} \frac{1}{24576}x^4e^{-x/4}\,dx \approx 0.747.$$

(On a TI83/84, fnInt($x^4 \times e^{-x/4} \div 24576, x, 0, 25$), where you type MATH 9:fnInt to get started.)

3. As in the previous part, the desired probability is

$$\int_{15}^{30} \frac{1}{24576} x^4 e^{-x/4}\, dx \approx 0.545.$$

Example 2 A certain type of saw blade used in a sawmill has a lifetime in hours that has a Weibull distribution with $\alpha = 2$ and $\beta = 10\,\text{h}$.

1. What is the probability that the blade fails in less than 8 hours of use?
2. What is the expected lifetime of this type of blade?

Solution 2.

1. The cdf is $F(x; \alpha, \beta) = 1 - e^{-(x/\beta)^\alpha}$ (if $x > 0$), so in the present case $F(8\,\text{h}; 2, 10\,\text{h}) = 1 - e^{-(8/10)^2} = 1 - e^{-16/25} = 0.473$.
2. $\int_0^\infty x \cdot F(x; 2, 10)\, dx = \int_0^\infty x \cdot \frac{2}{10^2} x^{2-1} e^{-(x/10)^2}\, dx = \int_0^\infty \frac{1}{50} x^2 e^{-x^2/100}\, dx = 8.86\,\text{h}$, where you can use your TI83/84's fnInt function for the last step.

Example 3

1. A die is rolled 50 times. What is the probability that the average of the 50 rolls is between 3.0 and 3.5?
2. A die is rolled 100 times. What is the probability that the average of the 100 rolls is between 3.0 and 3.5?

Solution 3. When rolling a die, the mean is $\mu = \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{7}{2}$, and the variance is $\sigma^2 = \sum_{i=1}^6 (i - \mu)^2 \cdot \frac{1}{6} = \frac{35}{12}$, so the standard deviation is $\sigma = \frac{\sqrt{105}}{6}$.

When rolling n dice the mean is then $\mu = \frac{7}{2}$ and the standard deviation is $\frac{\sigma}{\sqrt{n}}$.

You can now approximate with a normal distribution.

1. normalcdf$(3, 3.5, 7/2, \frac{\sqrt{105}}{6\sqrt{50}}) = 0.481$.

   You get a better approximation if you use the following continuity correction: normalcdf$(3 - \frac{1}{100}, 3.5 + \frac{1}{100}, 7/2, \frac{\sqrt{105}}{6\sqrt{50}}) = 0.499$.

2. Similarly, normalcdf$(3, 3.5, 7/2, \frac{\sqrt{105}}{6\sqrt{100}}) = 0.498$.

   With the continuity correction: normalcdf$(3 - \frac{1}{200}, 3.5 + \frac{1}{200}, 7/2, \frac{\sqrt{105}}{6\sqrt{100}}) = 0.510$.

Example 4 Suppose that the average annual income of all fully employed Canadian females between the ages of 30 and 40 is \$40 000, and that the standard deviation of these incomes is \$10 000. A random sample of 400 is drawn from this population.

1. What is the probability that the sample average is (a) at most \$41 000? (b) within \$800 of the population mean?

2. What value would the sample mean exceed 90% of the time?

3. What symmetric interval about the population mean would the sample mean fall in 75% of the time?

4. For 95% of the time, you would expect the population mean to be no more than a certain amount different from the sample mean. Find the amount.

Solution 4. Here $\mu = 40\,000$ and $\sigma = 10\,000/\sqrt{400} = 500$.

1. (a) $\text{normalcdf}(-\infty, 41\,000, 40\,000, 500) = 0.977$. (Whether you use $-\infty$, -10^{99}, or 0 as the left endpoint makes no difference to the first (at least) 500 digits of this probability.)
   (b) $\text{normalcdf}(39\,200, 40\,800, 40\,000, 500) = 0.890$.

2. Since $\text{invNorm}(0.10, 40\,000, 500) = \$39\,359.22$, the sample mean would be less than this amount 10% of the time and therefore exceed it 90% of the time.

3. Since
$$\text{invNorm}(0.125, 40\,000, 500) = \$39\,424.83$$
and
$$\text{invNorm}(0.875, 40\,000, 500) = \$40\,575.17,$$
the sample mean is in the interval $[\$39\,424.83, \$40\,575.17]$ 75% of the time. (Notice that this interval is symmetric about \$40 000 because the normal distribution is symmetric about the mean.)

4. Using the symmetry of the normal distribution again, the sample mean would be larger than $\text{invNorm}(0.975, 40\,000, 500) = \$40\,979.98$ 2.5% of the time (and smaller than $\text{invNorm}(0.025, 40\,000, 500) = \$39\,020.02$ another 2.5% of the time). Therefore the sample mean differs from the mean by at least $\$40\,979.98 - \$40\,000 = \$979.98$ 95% of the time.