Statistics Spring 2013 Test 2: Solutions

1. A random variable is distributed with the probability density $f(x) = k \cdot x$ for x from the interval $[0, 1]$ and $f(x) = 0$ for all other x. Find: (a) the value k, (b) the cumulative distribution function for x, (c) the expected value of x, and (d) the probability that x does not exceed 0.5.

Solution: From the condition that $\int f(x)\,dx = 1$, we conclude that $\int_0^1 f(x)\,dx = 1$. Substituting $f(x) = k \cdot x$ into this formula, we get

$$\int_0^1 k \cdot x\,dx = k \cdot \left.\frac{x^2}{2}\right|_0^1 = k \cdot \frac{1}{2} = \frac{k}{2}.$$

Thus, from the condition that

$$\frac{k}{2} = 1,$$

we conclude that $k = 2$. Thus, $f(x) = 2x$.

The cumulative distribution $F(x)$ can be obtained by using a formula

$$F(x) = \int_0^x f(t)\,dt = \int_0^x 2t\,dt = \left. x^2 \right|_0^x = x^2.$$

The expected value of x can be obtained by using a formula

$$\int x \cdot f(x)\,dx = \int_0^1 2x^2\,dx = \left.\frac{2x^3}{3}\right|_0^1 = \frac{2}{3}.$$

The probability that x does not exceed 0.5 is, by definition of the cumulative distribution function, equal to $F(0.5)$. We know that $F(x) = x^2$, so $F(0.5) = 0.25$.

2. A computer is connected to two printers. If at least one of them works, we can still print. The time-to-repair of each printer is exponentially distributed with the mean value of 1 year. What is the probability that after 1 year, we will still be able to print?

Solution: We cannot print is both printers are unavailable. The probability that one of the printers is available after time t is determined by the exponential distribution $F(t) = \exp(-\lambda \cdot t)$, so the the probability that it is not available is equal to $1 - F(t) = 1 - \exp(-\lambda \cdot t)$. The mean value of the exponential distribution is equal to $E = \frac{1}{\lambda}$, so from $E = 1$, we conclude that $\lambda = 1$. Thus, for $t = 1$, the probability that a printer is not working after 1 year is $1 - \exp(-1)$.

Since the printers are independent, the probability that both of them are not working is the product $(1 - \exp(-1))^2$. Thus, the probability that one of them is working is equal to $1 - (1 - \exp(-1))^2$.

3. For a printer-producing company, the probability that a printer needs repairs is 10%. In a sample of 100 printers, what is the expected value of number of faulty printers, and what is the standard deviation? What is the probability that no more than 15 printers are faulty (just give a formula, it is not necessary to get a numerical answer).

Solution: For each printer, the expected number of faulty printers is 0.1, and the variance of this number is $p \cdot (1 - p) = 0.1 \cdot 0.9 = 0.09$.

For the sum of n independent identically distributed random variables, both the mean and the variance are n times the mean and the variance of each variable. Thus, the mean is $100 \cdot 0.1 = 10$, an the variance is $100 \cdot 0.09 = 9$. The standard deviation is thus equal to the square root of the variance, i.e., we have $\sigma = 3$.

For the probability that no more than $z = 15$ printers are faulty, we can use the Central Limit Theorem, according to which this probability is equal to

$$\Phi\left(\frac{z - \mu}{\sigma}\right) = \Phi\left(\frac{15 - 10}{3}\right) = \Phi\left(\frac{5}{3}\right).$$

4. Two independent tasks are coming to the server. Their arrival times are uniformly distributed between noon and 5 pm. Compute the expected time of the task which arrives first, and the expected time of the task which arrives second.

Solution: Let t_1 be the arrival time of the first task and t_2 the arrival time of the second task. The arrival time of the latest task is then $\ell = \max(t_1, t_2)$. It is clear that for every value t, $\max(t_1, t_2) \le t$ if and only if $t_1 \le t$ and $t_2 \le t$. Since for t_1, the distribution is uniform, the probability

$$P(t_1 \le t) = \frac{t}{5}.$$

Similarly,

$$P(t_2 \le t) = \frac{t}{5}.$$

Since the two events are independent, the probability that both times are $\le t$ is equal to the product of these probabilities:

$$P(\ell \le t) = P(t_1 \le t \,\&\, t_2 \le t) = \left(\frac{t}{5}\right)^2 = \frac{t^2}{25}.$$

This is the cumulative distribution function $F_\ell(t)$ for the variable ℓ. The corresponding probability density function can be obtained as the derivative

$$f_\ell(t) = \frac{dF_\ell(t)}{dt} = \frac{d}{dt}\left(\frac{t^2}{25}\right) = \frac{2t}{25}.$$

Thus, the expected value of ℓ can be determined as

$$\int t \cdot f_\ell(t)\,dt = \int_0^5 \frac{2}{25} \cdot t^2\,dt = \frac{2}{25} \cdot \left.\frac{t^3}{3}\right|_0^5 = \frac{2}{25} \cdot \frac{25 \cdot 5}{3} = \frac{10}{3} = 3\frac{1}{3}.$$

So, the expected time of the task that arrives second is 3:20 pm.

The arrival time of the earlier task is $\min(t_1, t_2)$. It is easy to check that for every two numbers t_1 and t_2, the sum of $\min(t_1, t_2) + \max(t_1, t_2)$ consists of the sum of these same two numbers, but maybe in different order:

$$\min(t_1, t_2) + \max(t_1, t_2) = t_1 + t_2.$$

Thus,

$$\min(t_1, t_2) = t_1 + t_2 - \max(t_2, t_2).$$

Thus, for the expected values, we have

$$E[\min(t_1, t_2)] = E[t_1] + E[t_2] - E[\max(t_1, t_2)].$$

Since t_1 is uniformly distributed on the interval $[0, 5]$, its expected value is equal to

$$E[t_1] = \frac{0 + 5}{2} = \frac{5}{2}.$$

Similarly,

$$E[t_2] = \frac{5}{2}.$$

We already know that

$$E[\max(t_1, t_2)] = \frac{10}{3}.$$

Thus,

$$E[\min(t_1, t_2)] = \frac{5}{2} + \frac{5}{2} - \frac{10}{3} = \frac{5}{3} = 1\frac{2}{3}.$$

So, the expected time of the task that arrives first is 1:40 pm.

5. Derive a formula and explain how to generate a random variable with the density $f(x) = 0.5 \cdot x$ on the interval $[0, 2]$ if you have a standard random number generator U which produces values uniformly distributed on the interval $[0, 1]$. Use the inverse transform method.

Solution: According to the inverse transform method, we simulate the random variable as $F^{-1}(U)$, where $F(x)$ is the cumulative distribution function, and F^{-1} is the inverse function. Here,

$$F(x) = \int_0^x 0.5 \cdot t\, dt = 0.5 \cdot \left.\frac{t^2}{2}\right|_0^x = \frac{x^2}{4}.$$

The inverse function $F^{-1}(p)$ can be obtained from the fact that $x = F^{-1}(p)$ if and only if $F(x) = p$. In our case,

$$\frac{x^2}{4} = p$$

implies that $x^2 = 4 \cdot p$, hence $x = 2 \cdot \sqrt{p}$.

So, we generate the random variable as $2 \cdot \sqrt{U}$.

6. Suppose that we have a distribution in which we have 1 with probability θ and -1 with probability $1 - \theta$. Use method of moments and maximum likelihood to estimate the parameter θ from the following sample: -1, 1, -1, 1, -1, 1, -1, -1.

Solution: In the method of moments, we equate theoretical moments of the distribution with the sample moments. The number of resulting equations is equal to the number of moments that we consider. In general, we need as many equations as there are unknowns.

In our case, we have only one parameter, so it is sufficient to consider only one equation – i.e., only the first moment. Here, we have $n = 2$ possible values $x_1 = 1$ and $x_2 = -1$, with probabilities $p_1 = \theta$ and $p_2 = 1 - \theta$, so the theoretical expected value is equal to

$$\sum_{i=1}^{n} p_i \cdot x_i = \theta \cdot 1 + (1 - \theta) \cdot (-1) = 2\theta - 1.$$

The sample average is equal to

$$\frac{(-1)+1+(-1)+1+(-1)+1+(-1)+(-1)}{8} = -\frac{1}{4}.$$

By equating the theoretical and the sample moments, we get

$$2\theta - 1 = -\frac{1}{4},$$

hence

$$2\theta = 1 - \frac{1}{4} = \frac{3}{4}$$

and

$$\theta = \frac{3}{8}.$$

Let us now estimate θ by using the maximum likelihood method. The probability of -1 is equal to $1-\theta$, the probability of 1 is θ. Since different elements of the sample are independent, we take a product of these values:

$$(1-\theta)\cdot\theta\cdot(1-\theta)\cdot\theta\cdot(1-\theta)\cdot\theta\cdot(1-\theta)\cdot(1-\theta) = \theta^3\cdot(1-\theta)^5.$$

We want to find θ for which this likelihood value is the largest. As in the book, instead of maximizing the likelihood, we maximize its logarithm

$$\ln(\theta^3\cdot(1-\theta)^5) = 3\cdot\ln(\theta) + 5\cdot\ln(1-\theta).$$

Differentiating this expression with respect to θ and equating the derivative to 0, we conclude that

$$\frac{3}{\theta} - \frac{5}{1-\theta} = 0.$$

By bringing both fractions to the common denominator, we conclude that

$$3\cdot(1-\theta) - 5\theta = 0,$$

i.e., $3 - 8\theta = 0$ and

$$\theta = \frac{3}{8}.$$

7. In a shipment of 100 printers, 10 turns out to be faulty. Construct a 90% confidence interval for the proportion of faulty printers.

Solution: According to Section 9.3.2, the $(1-\alpha)\cdot 100\%$ confidence interval for a population proportion is

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}.$$

Here,

$$\hat{p} = \frac{10}{100} = 0.1.$$

To get $(1-\alpha)\cdot 100 = 90$, we take $\alpha = 0.1$. Thus,

$$\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{0.1\cdot 0.9}{100}} = \sqrt{\frac{0.09}{100}} = \frac{\sqrt{0.09}}{\sqrt{100}} = \frac{0.3}{10} = 0.03.$$

Hence, the confidence interval has the form

$$0.1 \pm z_{0.05} \cdot 0.03.$$