Homework 44

The **due date** for this homework is **Tue 7 May 2013 12:00 AM EDT**.

Question 1

Find the expectation $\mathbb{E}$ and variance $\mathbb{V}$ of x if its probability density function is $\rho(x) = (n+1)x^n$ (n a positive integer) on $[0, 1]$.

- $\mathbb{E} = 1$, $\mathbb{V} = \dfrac{n+2}{n+3}$.
- $\mathbb{E} = 1$, $\mathbb{V} = \dfrac{n+1}{n+2} - 1$.
- $\mathbb{E} = 1$, $\mathbb{V} = \dfrac{n+1}{n+2}$.
- $\mathbb{E} = \dfrac{n+1}{n+2}$, $\mathbb{V} = \dfrac{n+2}{n+3} - \left(\dfrac{n+1}{n+2}\right)^2$.
- $\mathbb{E} = \dfrac{n+1}{n+2}$, $\mathbb{V} = \dfrac{n+1}{n+3} - \left(\dfrac{n+1}{n+2}\right)^2$.
- $\mathbb{E} = \dfrac{n+1}{n+2}$, $\mathbb{V} = \dfrac{n+2}{n+3}$.

Question 2

Find the expectation $\mathbb{E}$ and variance $\mathbb{V}$ of x if its probability density function is $\rho(x) = \dfrac{2}{\pi}\dfrac{1}{x^2+1}$ on $[0, +\infty)$.

- $\mathbb{E} = \dfrac{2}{\pi}$, $\mathbb{V} = \dfrac{4}{\pi^2}$.
- $\mathbb{E} = \dfrac{2}{\pi}$, but $\mathbb{V}$ diverges.
- $\mathbb{E} = 1$, but $\mathbb{V}$ diverges.

- $\mathbb{E} = \dfrac{2}{\pi}, \mathbb{V} = \dfrac{2}{\pi} - \dfrac{4}{\pi^2}.$
- $\mathbb{E} = 1, \mathbb{V} = 1.$
- Both $\mathbb{E}$ and $\mathbb{V}$ diverge.

Question 3

Find the expectation $\mathbb{E}$ and variance $\mathbb{V}$ of n if its probability density function is $\rho(n) = \dfrac{1}{4}$ on $n = 1, 2, 3, 4.$

Hint: although we have not talked about expectation and variance for discrete probability distributions, you can do this! Think of the analogy with masses: expectation is center of mass and variance is moment of inertia. This problem hints at the fact that you can think of sums as discrete versions of integrals, opening the door to using Calculus in situations in which inputs are discrete but outputs are continuous. Much more about this in Chapter 5: Discretization!

- $\mathbb{E} = \dfrac{5}{2}, \mathbb{V} = \dfrac{7}{2}.$
- $\mathbb{E} = \dfrac{5}{2}, \mathbb{V} = \dfrac{5}{4}.$
- $\mathbb{E} = 2, \mathbb{V} = \dfrac{5}{4}.$
- $\mathbb{E} = \dfrac{5}{2}, \mathbb{V} = \dfrac{15}{4}.$
- $\mathbb{E} = 2, \mathbb{V} = \dfrac{7}{2}.$
- $\mathbb{E} = 2, \mathbb{V} = \dfrac{1}{2}.$

Question 4

The *median* m of a (one-dimensional) continuous probability distribution on $[a, b]$

is defined to be the value of x for which the probability of $x < m$ is equal to the probability of $x > m$ —that is, $1/2$. In the language of integrals, this is:

$$\int_a^m \rho(x)dx = \frac{1}{2} = \int_m^b \rho(x)dx$$

Find the value of the median for an exponential distribution with probability density function

$$\rho(x) = \alpha e^{-\alpha x} \qquad \text{on } [0, +\infty)$$

- $m = \dfrac{\ln 2}{\alpha}$
- $m = \dfrac{1}{\alpha}$
- $m = \alpha \ln 2$
- $m = \dfrac{1}{\alpha \ln 2}$
- $m = \dfrac{\alpha}{\ln 2}$
- $m = \alpha$

Question 5

There is a host of other numbers that one can associate to a probability distribution that generalize the median: e.g., the so-called *quantiles*. Let us consider an example —the *quartiles*:

- the first (or lower) quartile is the unique value Q_1 for which the probability of $x < Q_1$ is $1/4$;
- the second quartile (really the median) is the unique value Q_2 for which the probability of $x < Q_2$ is $2/4$;
- the third (or upper) quartile is the unique value Q_3 for which the probability of $x < Q_3$ is $3/4$;

You also have *quintiles*, *deciles*, the ubiquitous *percentiles*, etc.

Find the value of the first and third quartiles of the exponential distribution of the previous problem with $\rho(x) = \alpha e^{-\alpha x}$.

- $Q_1 = \dfrac{\ln(4/3)}{\alpha}, Q_3 = \dfrac{\ln 4}{\alpha}$.
- $Q_1 = \dfrac{\ln 4}{\alpha}, Q_3 = \dfrac{\ln(4/3)}{\alpha}$.
- $Q_1 = \dfrac{1}{\alpha \ln(4/3)}, Q_3 = \dfrac{1}{\alpha \ln 4}$.
- $Q_1 = \dfrac{1}{\alpha \ln 4}, Q_3 = \dfrac{1}{\alpha \ln(4/3)}$.
- $Q_1 = \dfrac{\alpha}{\ln(4/3)}, Q_3 = \dfrac{\alpha}{\ln 4}$.
- $Q_1 = \dfrac{\alpha}{\ln 4}, Q_3 = \dfrac{\alpha}{\ln(4/3)}$.

Question 6

Here is one last commonly used concept that captures some information about a probability distribution: the *mode* of a continuous probability distribution is the value M at which its probability density function has its maximum value.

Consider the so-called *beta* distribution, whose probability density function is given by

$$\rho(x) = \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} \qquad \text{on } [0,1]$$

Here the parameters α and β are positive real numbers , and $B(\alpha,\beta)$ is the constant necessary for $\rho(x)$ to integrate to 1 over the interval $[0,1]$. Compute the mode of this distribution under the assumptions $\alpha > 1$ and $\beta > 1$.

For the curious: although we have defined the mode as *the* value at which its probability density function achieves its absolute maximum, it might happen that there are two or more values at which this maximum is attained. It might also happen that there is no such value! The latter is the case when both $\alpha < 1$ or

$\beta < 1$.

- $M = \dfrac{\beta - 1}{\alpha + \beta - 2}$
- $M = \dfrac{\alpha - 1}{\beta - 1}$
- $M = \dfrac{\beta - 1}{\alpha - 1}$
- $M = \dfrac{\alpha + \beta - 2}{\alpha - 1}$
- $M = \dfrac{\alpha - 1}{\alpha + \beta - 2}$
- $M = \dfrac{\alpha + \beta - 2}{\beta - 1}$

Question 7

Suppose that a certain quantity follows a normal probability distribution with expectation μ and variance σ^2. That means that the probability that $a < x < b$ is computed by the integral

$$P(a < x < b) = \int_{x=a}^{b} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx \qquad (*)$$

In Lecture we mentioned three particular values of this probability:

- if $a = \mu - \sigma$ and $b = \mu + \sigma$, it is around 68%,
- if $a = \mu - 2\sigma$ and $b = \mu + \sigma$, it is around 95%, and
- if $a = \mu - 3\sigma$ and $b = \mu + 3\sigma$, it is around 99%.

But how can you compute other values? The trick consists of performing a simple substitution in the integral $(*)$:

$$y = \frac{x - \mu}{\sigma}$$

This yields

$$\int_{x=a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, dx = \int_{y=(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$

You can recognize the right hand side of this last equation as the probability of $\dfrac{a-\mu}{\sigma} < y < \dfrac{b-\mu}{\sigma}$, where y follows a normal distribution with expectation 0 and variance 1 —sometimes referred to as a *standard normal distribution*. Recall from Question 6 of Homework 43 that the latter can be expressed as the difference of the cumulative distribution function F associated to y as

$$F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

Now, this cumulative distribution function is usually written in terms of a function called the *error function*, erf, as

$$F(z) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)$$

Putting everything together, we arrive at the formula

$$P(a < x < b) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{b-\mu}{\sigma\sqrt{2}}\right) - \mathrm{erf}\left(\frac{a-\mu}{\sigma\sqrt{2}}\right)\right]$$

Before the age of computers, the values of this error function were tabulated for reference. Nowadays, you can use WolframAlpha or any other computer algebra system to compute its values.

According to the 2008 National Health Statistics Report of the U.S. Department of Health and Human Services, the expected value of the height for adult males (age 20 and above) in the U.S. is $176.3\,\mathrm{cm}$, with a standard deviation of $8.0\,\mathrm{cm}$. Use WolframAlpha or your computer algebra system of choice to estimate the probability that an adult American male is between $170\,\mathrm{cm}$ and $180\,\mathrm{cm}$ tall.

[**Reference:** 2008 National Health Statistics Report, U.S. Department of Health and Human Services]

- 43%
- 46%

- 52%
- 55%
- 40%
- 49%

In accordance with the Honor Code, I certify that my answers here are my own work.

Submit Answers Save Answers