Weekly Notes for EE2012 2013/14 – Week 8

T. J. Lim

March 10, 2015

Book sections covered this week: 4.3 – 4.4.

1 Expected Values

1.1 Definition

For a discrete random variable X, the expected value was defined in an earlier class as

$$E[X] = \sum_k x_k p_X(x_k)$$

where $p_X(x_k)$, $x_k \in S_X$, is the PMF of X. For a repeatable experiment, this expression could be interpreted as the large-N limit of the arithmetic average of $x(1), x(2), \dots, x(N)$ where $x(i)$ is the value of X obtained in the i-th repetition of the experiment, i.e.

$$\lim_{N\to\infty} \langle X \rangle_N = E[X], \tag{1}$$

where $\langle X \rangle_N = \frac{1}{N}\sum_{i=1}^{N} x(i)$.

The approach of defining the arithmetic mean of N samples of X, when N is large, as the expected value of X has to be more carefully carried out in the case of continuous X. First, suppose S_X has finite support[1]. We can divide S_X into N equal-width non-overlapping segments, each with width Δ, and denote the k-th segment by $(x_k, x_{k+1}]$, $k = 0, \dots, N-1$, with $x_{k+1} - x_k = \Delta$.

Since $\{(x_k, x_{k+1}]\}_{k=0}^{N-1}$ partition S_X, we have

$$\langle X \rangle_N \approx \sum_{k=0}^{N-1} \frac{x_k N_k}{N} = \sum_{k=0}^{N-1} x_k r_k(N) \tag{2}$$

where N_k is the number of occurrences of $\{X \in (x_k, x_{k+1}]\}$ in N trials, and $r_k(N) = N_k/N$ is its relative frequency.

Now, the probability of $\{X \in (x_k, x_{k+1}]\}$ when Δ is small will be, by definition,

$$P[x_k < X \le x_{k+1}] \approx f_X(x_k)\Delta, \tag{3}$$

[1] Meaning that the total length of S_X is finite, e.g. $S_X = [0,1) \cup (2,3]$ has finite support, whereas $S_X = [0,\infty)$ does not.

and also, $\lim_{N\to\infty} r_k(N) = P[x_k < X \leq x_{k+1}]$. Therefore, (2) and (3) together imply that

$$\lim_{N\to\infty} \langle X \rangle_N = \lim_{N\to\infty} \sum_{k=0}^{N-1} x_k f_X(x_k)\Delta = \int_{S_X} x f_X(x) dx. \tag{4}$$

Next, we extend the result to all S_X, with or without finite support, by integrating over all real values. Finally, we take our final expression to be a definition of $E[X]$ whether or not the experiment is repeatable:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{5}$$

Note that this expression holds for *all types* of random variables.

Example 1: X is a uniform continuous random variable, with PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

Its mean value is

$$\begin{aligned} E[X] &= \int_a^b \frac{x}{b-a} dx & (6) \\ &= \frac{1}{2(b-a)} \left. x^2 \right|_a^b & (7) \\ &= \frac{1}{2(b-a)} (b^2 - a^2) & (8) \\ &= \frac{a+b}{2} & (9) \end{aligned}$$

where the last line results from the identity $b^2 - a^2 = (b-a)(b+a)$. ■

Example 2: Suppose the PDF of X is

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0 \tag{10}$$

where $\lambda > 0$ is a constant. Then $E[X] = \int_0^\infty \lambda x e^{-\lambda x} dx$, which has to be evaluated using integration by parts with the substitutions:

$$\begin{aligned} u(x) = x \qquad & \frac{dv(x)}{dx} = \lambda e^{-\lambda x} & (11) \\ \frac{du(x)}{dx} = 1 \qquad & v(x) = -e^{-\lambda x}. & (12) \end{aligned}$$

We then obtain

$$\begin{aligned} E[X] &= \left. -x e^{-\lambda x} \right|_0^\infty + \int_0^\infty e^{-\lambda x} dx & (13) \\ &= \frac{1}{\lambda}. & (14) \end{aligned}$$

1.2 Helpful Tricks

Computing the mean value using its definition often requires rather heroic feats of integration. Here are some useful tricks that may be helpful in finding expected values.

1. *If $f_X(x)$ has even symmetry about $x = m$, then $E[X] = m$.* "Even symmetry" means that $f_X(m-a) = f_X(m+a)$ for any $a > 0$. The proof of this result is left as an exercise.

2. *If X is a non-negative random variable, then $E[X] = \int_0^\infty 1 - F_X(t)dt$.* The proof can be obtained through a graphical view of the definition of $E[X]$, keeping in mind the non-negativity of X.

3. *If X is a non-negative integer-valued random variable, then $E[X] = \sum_{k=0}^\infty P[X > k]$.* This follows directly from the previous trick, if we recall that $F_X(x)$ for a discrete r.v. X is a staircase function.

Example 3: The Gaussian PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{15}$$

has even symmetry around $x = \mu$. Therefore, by Trick 1 above, $E[X] = \mu$. ■

Example 4: For the r.v. of Example 2, we have

$$F_X(t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}, \quad t > 0.$$

Using Trick 2, we have

$$\begin{aligned} E[X] &= \int_0^\infty e^{-\lambda t} dt & (16) \\ &= \frac{1}{\lambda} e^{-\lambda t}\Big|_\infty^0 & (17) \\ &= \frac{1}{\lambda} & (18) \end{aligned}$$

In contrast to Example 2, we do not need to use integration by parts to find the mean of X now. ■

1.3 Mean of $g(X)$

Similar to the discrete r.v. case, we have for any random variable,

$$E[g(X)] = \int_{-\infty}^\infty g(x) f_X(x) dx. \tag{19}$$

The above definition with $g(x) = \sum_{k=1}^{n} g_k(x)$ leads to

$$E\left[\sum_{k=1}^{n} g_k(X)\right] = \int_{-\infty}^{\infty} \sum_{k=1}^{n} g_k(x) f_X(x) dx \tag{20}$$

$$= \sum_{k=1}^{n} \int_{-\infty}^{\infty} g_k(x) f_X(x) dx \tag{21}$$

$$= \sum_{k=1}^{n} E[g_k(X)]. \tag{22}$$

When $g(x) = c$, a constant, then we see that

$$E[c] = \int_{-\infty}^{\infty} c f_X(x) dx = c.$$

When $g(x) = cX$, then $E[cX] = cE[X]$. These results tell us that the expectation operator $E[\cdot]$ is *linear*, in that for any set of n functions $g_i(x)$ and constants c_i, $i = 1, \ldots, n$, we have

$$E[c_1 g_1(X) + \cdots c_n g_n(X)] = c_1 E[g_1(X)] + \cdots + c_n E[g_n(X)]. \tag{23}$$

1.4 Variance of X

For the special case of $g(X) = (X - \mu)^2$ where $\mu = E[X]$, $E[g(X)]$ is the variance of X, i.e.

$$\mathrm{var}(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx. \tag{24}$$

It is also straightforward to show that $\mathrm{var}(X) = E[X^2] - \mu^2$.

Example 5: Find the variance of the uniform random variable of Example 1.

$$\sigma_X^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx \tag{25}$$

$$= \frac{1}{b-a} \int_{(a-b)/2}^{(b-a)/2} y^2 dy \quad (\text{Let } y = x - \mu) \tag{26}$$

$$= \frac{(b-a)^2}{12}. \tag{27}$$

This is a famous result that has been applied to the analysis of quantization noise.

2 Important Continuous Random Variables

2.1 Uniform

The mean and variance of the uniform random variable were defined earlier. This distribution is used to model a random variable that can take any real value in (a, b), and has equal probability of being in any δx-neighbourhood in that range.

2.2 Exponential

Consider a Poisson arrival process, i.e. events that occur randomly in time and independently from each other, such as arrivals of page requests at a web server, or cars at a parking lot, etc. The average rate of arrival is known, and denoted by λ (arrivals per unit time), but nothing else.

Let T represent the waiting time for the next arrival (or Poisson point). What is the distribution of T? To answer this question, we first define $N(t)$ as the number of arrivals between the present (defined as time 0) and time t, where t can be any positive value. Given that the rate of arrivals is λ, the average number of arrivals in t time units must be $\alpha = \lambda t$. Therefore, the PMF of $N(t)$ is

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \ldots \tag{28}$$

Now consider the event $\{T > t\}$, or in words, the waiting time for the next arrival exceeds t. This event occurs if and only if $N(t) = 0$, because if no events occur in t seconds, the waiting time for the next arrival must exceed t seconds; and if the waiting time for the next arrival exceeds t seconds, then there must have been no arrivals in that time. Therefore, $\{T > t\}$ is equivalent to $\{N(t) = 0\}$, and hence

$$1 - F_T(t) = P[T > t] = P[N(t) = 0] = e^{-\lambda t}. \tag{29}$$

The CDF of T is thus $F_T(t) = 1 - e^{-\lambda t}$, $t > 0$. For $t \leq 0$, then $F_T(t) = 0$ since waiting time cannot be negative. Thus the PDF of T is

$$f_T(t) = \frac{d}{dt} F_T(t) = \lambda e^{-\lambda t}, \quad t > 0. \tag{30}$$

This is the PDF of the exponential random variable, which models the inter-arrival times of a Poisson arrival process.

The mean of T was shown in Examples 2 and 4 to be $1/\lambda$. The variance can be shown to be $\mathrm{var}(T) = 1/\lambda^2$.

Memoryless Property The exponential random variable is the only continuous random variable to possess the memoryless property:

$$P[T > t + h | T > t] = P[T > h]. \tag{31}$$

The proof of this result is straightforward, and left as an exercise. This property tells us that no matter how long one has already waited for the next Poisson point to arrive, the probability of waiting another h time units is as if one only started waiting.

For instance, if the appearance of a rare bird in your backyard (just pretend you have one) can be modelled as a Poisson process, then it doesn't matter if you have already spent 2 days in a vigil waiting for one to appear. You will have the same chance of seeing one in the next five minutes as you would if you had not been keeping watch for two days.

2.3 Gaussian or Normal

The Gaussian distribution occupies a very special place within probability and statistics, due to several reasons:

1. It is (surprisingly) easy to manipulate e.g. if X is Gaussian, then $aX + b$ is also Gaussian, it is symmetric about its mean value, its CDF can be written in terms of the CDF of a "standard normal" random variable, etc.

2. It is the distribution that the sum of a large number of random variables, whatever their individual distributions, converges to under certain loose conditions. This is known as the **Central Limit Theorem (CLT)** and we may have time to study it at the end of the course[2].

Statisticians, who usually deal with data collected from a large number of sources (e.g. survey data), find the Gaussian r.v. so useful in their work that they call it the "normal" r.v.

The PDF of a Gaussian random variable X is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}, \tag{32}$$

where μ and σ are the two parameters of the distribution. It can be shown that $E[X] = \mu$ (since the PDF is symmetric around $x = \mu$) and $\mathrm{var}(X) = \sigma^2$ (not straightforward, see Example 4.19 in the textbook), and that $f_X(x)$ integrates to one (see Example 4.21 in the book).

As a notational shorthand, we can write $X \sim \mathcal{N}(\mu, \sigma^2)$ to indicate that X is Gaussian with a mean of μ and variance of σ^2. For instance, an $\mathcal{N}(0, 1)$ random variable is Gaussian with mean 0 and variance 1, and is known as the **standard normal** random variable.

2.3.1 Calculating Gaussian Probabilities

As there is no closed-form expression for $\int_{-\infty}^{x} \exp(-t^2)dt$, we cannot find one for the CDF of an $\mathcal{N}(\mu, \sigma^2)$ r.v. However, the following development reveals that probabilities involving a Gaussian r.v. with any mean and variance can be expressed in terms of the CDF of a standard normal $\mathcal{N}(0, 1)$ r.v.

The CDF of X is

$$F_X(x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-\mu)^2/2\sigma^2} dt. \tag{33}$$

By a change of variables $s = (t - \mu)/\sigma$, we have

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-s^2/2} ds. \tag{34}$$

[2] Briefly, if $X_1, X_2, \ldots, X_N$ are independent identically distributed (i.i.d.) random variables, then $\sum_{i=1}^{N} X_i \to \mathcal{N}(\mu, \sigma^2)$ as $N \to \infty$, where $\mu = \sum_{i=1}^{N} \mu_i$ and $\sigma^2 = \sum_{i=1}^{N} \sigma_i^2$, μ_i and σ_i^2 being the mean and variance of X_i, respectively.

But this is the CDF of an $\mathcal{N}(0,1)$ r.v. evaluated at $(x-\mu)/\sigma$, i.e.

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{35}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp(-t^2/2)dt$. Therefore, the CDF of an $\mathcal{N}(\mu,\sigma^2)$ r.v. can always be expressed in terms of that of an $\mathcal{N}(0,1)$ r.v., which means that a look-up table of values of $\Phi(z)$ is sufficient to handle *all* Gaussian probabilities.

A popular alternative to the $\Phi(z)$ function is the Q function, defined as the complementary CDF of the $\mathcal{N}(0,1)$ r.v., i.e.

$$Q(x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty} e^{-t^2/2}dt. \tag{36}$$

It is easily demonstrated using the integration by substitution method above that if $X \sim \mathcal{N}(\mu,\sigma^2)$, then

$$P[X > x] = Q\left(\frac{x-\mu}{\sigma}\right) = P\left[Z > \frac{x-\mu}{\sigma}\right] \tag{37}$$

where $Z \sim \mathcal{N}(0,1)$.

Example 6: Let $X \sim \mathcal{N}(2,4)$, i.e. X is Gaussian with $E[X] = 2$ and $\mathrm{var}(X) = 4$. Find the following probabilities:

1. $P[X > 5]$;
2. $P[0 < X \leq 3]$;
3. $P[X \leq 0]$

in terms of the Q function.
Answers: We have

$$P[X > 5] = Q\left(\frac{5-2}{2}\right) = Q(1.5). \tag{38}$$

To find the second probability, we note that $P[0 < X \leq 3] = P[X > 0] - P[X > 3]$, and therefore

$$P[0 < X \leq 3] = Q\left(\frac{0-2}{2}\right) - Q\left(\frac{3-2}{2}\right) = Q(-1) - Q(0.5). \tag{39}$$

Due to the symmetry of the $\mathcal{N}(0,1)$ PDF about $x = 0$, we have $Q(x) = 1 - Q(-x)$, and hence there is no need to tabulate values of $Q(x)$ for $x < 0$. We therefore always express our answers in terms of Q functions with argument values that are non-negative. For this case, we substitute $Q(-1) = 1 - Q(1)$, and obtain

$$P[0 < X \leq 3] = 1 - Q(1) - Q(0.5). \tag{40}$$

Finally,

$$P[X \leq 0] = 1 - P[X > 0] = 1 - Q(-1) = Q(1). \tag{41}$$

From Table 4.2, we can check that $P[X > 5] = 6.68 \times 10^{-2}$, $P[0 < X \leq 3] = 0.532$, and $P[X \leq 0] = 0.159$. ■

Example 7: Let X be an $\mathcal{N}(\mu, \sigma^2)$ random variable. Find the probability that $|X - \mu| > n\sigma$ for positive integer-valued n.

Answer: We have that

$$\begin{aligned} P[|X - \mu| > n\sigma] &= P[X - \mu < -n\sigma] + P[X - \mu > n\sigma] & (42) \\ &= 1 - P[X \geq \mu - n\sigma] + P[X > \mu + n\sigma] & (43) \\ &= 1 - Q(-n) + Q(n) & (44) \\ &= 2Q(n). & (45) \end{aligned}$$

A table of these probabilities is shown here:

n	1	2	3	4
$2Q(n)$	0.318	4.56×10^{-2}	2.70×10^{-3}	6.34×10^{-5}

It is clear that the probability of a Gaussian r.v. taking a value more than $n\sigma$ away from its mean is very low once n exceeds about 3. For instance if we model the average marks of a random student over a semester as Gaussian with mean 60 and standard deviation 10, then only about 1 in 185 students (1/0.0027) will score higher than 90 ($= \mu + 3\sigma$) or lower than 30 ($= \mu - 3\sigma$). ■

It is important to note that:

1. $Q(x)$ is a strictly decreasing function of x and decays very quickly (faster than e^{-x^2});
2. $Q(0) = 0.5$, and $Q(x) = 1 - Q(-x)$, due to the symmetry of the $\mathcal{N}(0, 1)$ PDF;
3. $Q(x)$ for $x > 0$ is the area under the "tail" of the $\mathcal{N}(0, 1)$ PDF to the right of x.

3 Diagnostic Questions

1. The average arrival rate of a Poisson process is 10 arrivals per hour. The inter-arrival time T is measured in minutes. What is the PDF of T?
2. Conditioned on A, $X \sim \mathcal{N}(0, 1)$ while conditioned on A^c, $X \sim \mathcal{N}(2, 1)$. Find $P[X > 1]$ in terms of the Q function if $P[A] = p_A$.