ECE 361 Probability for Engineers (Fall, 2016)
Lecture 3a

---

§2.3 Functions of random variables

Functions of RVs. A function of a RV X, say $Y = g(X)$, is a new RV.

Example. Let X be the temperature in degrees Celsius, and let $Y = 1.8X + 32$ be the temperature in degrees Fahrenheit. Then Y is a (in this case linear) function of X and thus Y is also an RV.

Let $\mathcal{X}$ be the support of the RV X and let $\mathcal{Y}$ be the support of the RV Y, defined as the set of all possible values taken by the function g. We find the PMF for Y from the PMF for X in the same way as we find the PMF for X from the random experiment $(\Omega, \mathcal{F}, \mathbb{P})$ – we look at the partition induced by the mapping. Namely, for each $y \in \mathcal{Y}$ we find the set $\mathcal{X}_y = \{x \in \mathcal{X} : g(x) = y\}$. The collection of sets $(\mathcal{X}_y, y \in \mathcal{Y})$ partitions $\mathcal{X}$, and we find the PMF for Y as

$$p_Y(y) = \sum_{x \in \mathcal{X}_y} p_X(x), \; y \in \mathcal{Y}. \tag{1}$$

Example. Let X be a RV with PMF that is uniform over $\mathcal{X} = \{-4, \ldots, 4\}$. Note $|\mathcal{X}| = 9$ and thus uniformity means $p_X(x) = 1/9$ for each $x \in \mathcal{X}$. Define $Y = |X|$ and note $\mathcal{Y} = \{0, 1, 2, 3, 4\}$. The preimages of each $y \in \mathcal{Y}$ are $\mathcal{X}_y = \{-y, y\}$, for $y \in \mathcal{Y}$ where of course $-0 = 0$. Then immediately we see

$$p_Y(0) = 1/9, \; p_Y(1) = p_Y(2) = p_Y(3) = p_Y(4) = 2/9. \tag{2}$$

Example. Same as above but define $Y = X^2$. Then $\mathcal{Y} = \{0, 1, 4, 9, 16\}$ and

$$p_Y(0) = 1/9, \; p_Y(1) = p_Y(4) = p_Y(9) = p_Y(16) = 2/9. \tag{3}$$

§2.4 Expectation, mean, and variance

The PMF fully characterizes the RV, in fact it defines the RV. The expectation is a summary of the PMF – it is a weighted sum of the values taken by the RV and the probabilities of those values.

A motivation for the expectation is as follows. Consider a wheel with n values $m_1, \ldots, m_n$ where each spin pays the corresponding value. You spin the wheel k times and these k spins result in $k_1, \ldots, k_n$ occurrences of each value. Your total payout is $k_1 m_1 + \cdots + k_n m_n$, and the payout per spin is

$$M(k) = \frac{k_1 m_1 + \cdots + k_n m_n}{k} = \frac{k_1}{k} m_1 + \cdots \frac{k_n}{k} m_n \tag{4}$$

As k grows large we expect the fractions k_i/k to converge to a value that we call the probability of value i, say $\lim_{k \to \infty} k_i/k = p_i$, for each $i \in [n]$. Then we see

$$\lim_{k \to \infty} M(k) = p_1 m_1 + \cdots p_n m_n. \tag{5}$$

That is, the asymptotic (in k) pay per spin is the weighted sum of the probabilities p_i times the payoffs m_i, summed over the values $i \in [n]$. This is the expectation.

Expectation. The expected value (also called the expectation and the mean) of the RV X with pmf $\mathbf{p}$ is the real number

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p(x). \tag{6}$$

The expectation is a **number**. It is not a random variable. Always[1].

Example. Consider two independent coin tosses, each with 3/4 probability of heads, and let X be the number of heads. This is a binomial RV with parameters $n = 2$ and $p = 3/4$. Its PMF is

$$p(k) = \begin{cases} (1/4)^2, & k = 0 \\ 2(1/4)(3/4), & k = 1 \\ (3/4)^2, & k = 2 \end{cases} . \tag{7}$$

Its expectation is

$$\mathbb{E}[X] = 0 \cdot \left(\frac{1}{4}\right)^2 + 1 \cdot 2 \left(\frac{1}{4}\right)\left(\frac{3}{4}\right) + 2 \cdot \left(\frac{3}{4}\right)^2 = \frac{24}{16} = \frac{3}{2}. \tag{8}$$

The interpretation of 3/2 is: if you were to play a game where you earn a dollar for each head when two coins are tossed, and you play this game many times, then 3/2 is the average payout per play. This is the "fair" price to play.

Another interpretation of $\mathbb{E}[X]$ is as the center of gravity of the PMF $\mathbf{p}$. Namely, if you assign a "mass" of $p(x)$ at location x to each $x \in \mathcal{X}$, then the value $\mathbb{E}[X]$ is the unique point at which you can balance this collection of weights with your finger. See Fig. 2.7.

Variance, moments, and the expected value rule

More generally, we define the nth moment of the RV X as $\mathbb{E}[X^n]$, i.e., the expected value of the random variable X^n. Note the expected value is simply the first moment.

The second most important number associated with a PMF, besides the mean, is the variance, defined as:

$$\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]. \tag{9}$$

The variance is the expected value of the RV $(X - \mathbb{E}[X])^2$, the squared amount by which the RV X differs from the mean. Note that $\mathrm{var}(X) \geq 0$ since the RV $(X - \mathbb{E}[X])^2$ is non-negative.

Closely related to the variance is its square root, which is the standard deviation:

$$\mathrm{std}(X) = \sqrt{\mathrm{var}(X)}. \tag{10}$$

Again, $\mathrm{std}(X) \geq 0$. If X is measured in units (say, meters, kilograms, seconds, etc.), then the variance has units squared (meters squared, kilograms squared, seconds squared, etc.), whereas the standard deviation has the same units as X. For this reason the standard deviation is sometimes more informative as an indicator of the expected "spread" around the mean value.

Example. Recall X uniform over $\{-4, \ldots, 4\}$ with $p_X(x) = 1/9$ for each x. Clearly (why?) $\mathbb{E}[X] = 0$. Define $Z = (X - \mathbb{E}[X])^2$ with support $\mathcal{Z} = \{0, 1, 4, 9, 16\}$ and PMF $p_Z(0) = 1/9$ and $p_Z(1) = p_Z(4) = p_Z(9) = p_Z(16) = 2/9$. Then

$$\mathrm{var}(X) = \mathbb{E}[Z] = 0 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} + 4 \cdot \frac{2}{9} + 9 \cdot \frac{2}{9} + 16 \cdot \frac{2}{9} = \frac{60}{9}. \tag{11}$$

Note that to find the variance above we first defined the RV $Z = (X - \mathbb{E}[X])^2$, found its PMF $\mathbf{p}_Z$ from the PMF $\mathbf{p}_X$, then computed $\mathbb{E}[Z]$ from $\mathbf{p}_Z$. There is an easier way. In general, to find the expected value of a function of an RV, say $Z = g(X)$, we don't need to find the PMF $\mathbf{p}_Z$.

Expectation of a function of an RV. Let X be an RV with PMF $\mathbf{p}_X$, and let $g(X)$ be an RV defined as a function g of X. Then $g(X)$ has expected value

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) p_X(x). \tag{12}$$

Note the above formula does not require $\mathbf{p}_Z$, it only requires $\mathbf{p}_X$. The proof is found in the book on page 85.

[1] Except for conditional expectation, but we aren't discussing that.

Example. For X uniform as above, we find

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \mathcal{X}} p_X(x)(x - \mathbb{E}[X])^2 = \frac{1}{9} \sum_{x=-4}^{4} x^2 = \frac{60}{9}. \tag{13}$$

A RV has a variance that is non-negative. The variance will equal zero only when $p_X(x)(x - \mathbb{E}[X])^2 = 0$ for each $x \in \mathcal{X}$. But this means $x = \mathbb{E}[X]$ for each $x \in \mathcal{X}$, which is to say the RV is trivial in that it only takes on one possible value, i.e., its support is a single point $\mathcal{X} = \{x\}$. In short, the variance equals zero when the RV is actually a constant. We summarize below.

Variance. The variance $\text{var}(X)$ of a RV X is defined by

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2], \tag{14}$$

and can be calculated as

$$\text{var}(X) = \sum_{x \in \mathcal{X}} p_X(x)(x - \mathbb{E}[X])^2. \tag{15}$$

It is always nonnegative. Its square root is the standard devation $\text{std}(X)$.

Properties of mean and variance

Linear functions of RVs enjoy several nice properties for easy computation of their mean and variance. Let $Y = aX + b$, i.e., we specify a linear function $g(x) = ax + b$ and use it to define a RV $Y = g(X)$. We first consider the mean:

$$\mathbb{E}[aX + b] = \sum_{x \in \mathcal{X}} p(x)(ax + b) = a \sum_{x \in \mathcal{X}} p(x)x + b \sum_{x \in \mathcal{X}} p(x) = a\mathbb{E}[X] + b. \tag{16}$$

Thus if $Y = aX + b$ then $\mathbb{E}[Y] = a\mathbb{E}[X] + b$. We call this property *linearity of expectation.* Next consider the variance:

$$\text{var}(Y) = \sum_{x \in \mathcal{X}} p(x)(ax + b - \mathbb{E}[aX + b])^2 = a^2 \sum_{x \in \mathcal{X}} p(x)(x - \mathbb{E}[X])^2 = a^2 \text{var}(X). \tag{17}$$

Thus if $Y = aX + b$ then $\text{var}(Y) = a^2\text{var}(X)$. This expression is more transparent when you view aX as a scaling of X and $X + b$ as a translation of X, and $\text{var}(X)$ as the "spread" around the mean. Clearly translation of X by b will not affect the spread, thus $\text{var}(X + b) = \text{var}(X)$. Further, scaling of X by a should scale the standard deviation by a. Thus $\text{std}(aX) = |a|\text{std}(X)$.

Mean and variance of a linear function of an RV. Let X be a RV and let $Y = aX + b$ where a, b are given scalars. Then

$$\mathbb{E}[Y] = a\mathbb{E}[X] + b, \ \text{var}(Y) = a^2\text{var}(X). \tag{18}$$

Finally, we give an even easier expression for computing the variance:

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{19}$$

Proof:

$$\begin{aligned}
\text{var}(X) &= \sum_{x \in \mathcal{X}} (x - \mathbb{E}[X])^2 p(x) \\
&= \sum_{x \in \mathcal{X}} (x^2 - 2x\mathbb{E}[X] + \mathbb{E}[X]^2) p(x) \\
&= \sum_{x \in \mathcal{X}} x^2 p(x) - 2\mathbb{E}[X] \sum_{x \in \mathcal{X}} x p(x) + \mathbb{E}[X]^2 \sum_{x \in \mathcal{X}} p(x) \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2.
\end{aligned} \tag{20}$$

Example. Average speed vs. average time. If the weather is good (which happens with probability 0.6), Alice walks the 2 miles to class at a speed of $V = 5$ miles per hour, and otherwise rides her motorcycle at a speed of $V = 30$ miles per hour. What is the mean of the time T to get to class? Define the RV T as the time to get to class with PMF:

$$p_T(t) = \begin{cases} 0.6, & t = 2/5 \\ 0.4, & t = 2/30 \end{cases}, \tag{21}$$

with mean

$$\mathbb{E}[T] = 0.6 \cdot \frac{2}{5} + 0.4 \cdot \frac{2}{30} = \frac{4}{15}. \tag{22}$$

Mean and variance of some common RVs

Bernoulli. Let $X \sim \text{Ber}(p)$ where $\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0)$. Then:

$$\begin{aligned} \mathbb{E}[X] &= 1 \cdot p + 0 \cdot (1-p) = p \\ \mathbb{E}[X^2] &= 1^2 \cdot p + 0^2 \cdot (1-p) = p \\ \text{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p). \end{aligned} \tag{23}$$

Note the variance as a function of p is zero for $p = 0$ or $p = 1$ and attains its maximum value of $1/4$ at $p = 1/2$ – why is this intuitive?

Discrete uniform RV. What is the mean and variance of the RV associated with the roll of a fair six-sided die? Here $\mathcal{X} = [6]$ and $p(x) = 1/6$ for $x \in \mathcal{X}$, and $\mathbb{E}[X] = 3.5$. Further:

$$\text{var}(X) = \frac{1}{6}(1^2 + \cdots + 6^2) - (3.5)^2 = \frac{35}{12}. \tag{24}$$

More generally, we define a discrete uniform RV taking values in the set $\mathcal{X} = \{a, a+1, \ldots, b-1, b\}$ for $a < b$ (note $|\mathcal{X}| = b - a + 1$) as having PMF $p(x) = 1/|\mathcal{X}| = 1/(b-a+1)$ for $x \in \mathcal{X}$. Then the mean is $\mathbb{E}[X] = (a+b)/2$. In the case where $a = 1$ and $b = n$ so $\mathcal{X} = [n]$ we find[2]

$$\mathbb{E}[X^2] = \frac{1}{n} \sum_{k=1}^{n} k^2 = \frac{1}{6}(n+1)(2n+1). \tag{25}$$

This allows us to find the variance as

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 = \cdots = \frac{n^2-1}{12}. \tag{26}$$

Now, using the fact that the variance is not affected by a shift, note that a general uniform RV with support $\mathcal{X} = \{a, a+1, \cdots, b-1, b\}$ has the same variance as $\mathcal{X} = \{1, \ldots, n\}$ for $n = b - a + 1$. Thus:

$$\text{var}(X) = \frac{(b-a+1)^2 - 1}{12} = \frac{(b-a)(b-a+2)}{12}. \tag{27}$$

References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.

[2] using the sum of squares equation: $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$.