Indian Institute of Technology Bombay
Department of Electrical Engineering

Handout 14 EE 325 Probability and Random Processes
Lecture Notes 9 August 28, 2014

1 Variance of a Random Variable

The expectation $\mathbb{E}[X]$ is also known as the first moment or mean of the random variable. Can we put some quantitative measure on how close the random variable is to its mean. In electrical engineering such situations frequently occur in sampling and quantization, where one wishes to know the quantization error around a level of finite bit representation. To put it more formally, we are trying to find the mean square error, i.e, the difference $X - a$ is squared and then expectation is taken to obtain $\mathbb{E}\left(X - a\right)^2$, where a is some value of interest, for example the quantized value after sampling. The variance of a random variable $Var(X)$ is defined as

$$Var(X) = \mathbb{E}\left(X - \mathbb{E}[X]\right)^2$$

In the notes, we will often denote $Var(X)$ by σ_X^2. A simple expansion of the RHS above results in the following.

Proposition 1

$$\mathbb{E}\left(X - \mathbb{E}[X]\right)^2 = \mathbb{E}X^2 - \left(\mathbb{E}[X]\right)^2.$$

The mean $\mathbb{E}X$ is the point of minimum mean squared error, as the following proposition states.

Proposition 2

$$\mathbb{E}(X - \mu)^2 \le \mathbb{E}(X - a)^2, \forall a \in \mathbb{R}$$

Proof: Let us maximize the RHS with respect to a.

$$\frac{\mathrm{d}}{\mathrm{d}a}\mathbb{E}(X - a)^2 = -2\mathbb{E}(X - a) = -2(\mathbb{E}X - a).$$

Thus the only optimum is at $a = \mathbb{E}X$, and since the double derivative is 2 everywhere, we have the desired result.

■

Let us now introduce the Poisson distribution and compute its variance as an example.

1.1 Poisson Distribution: $Poisson(\lambda)$

The Poisson distribution is over $\mathbb{N}$, and has a single parameter, namely $\lambda \in \mathbb{R}^+$. The distribution is

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Exercise 1 *Verify that the given distribution satisfies the axioms of probability.*

Let us compute the mean and variance of $Poisson(\lambda)$.

$$\begin{aligned}
\mathbb{E}(X) &= e^{-\lambda}\sum_{k\geq 0} k\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k\geq 1} k\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k\geq 1} \lambda\frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda}\lambda e^{\lambda} \\
&= \lambda.
\end{aligned}$$

Let us now compute $\mathbb{E}X^2 - \mathbb{E}X$.

$$\begin{aligned}
\mathbb{E}(X) &= e^{-\lambda}\sum_{k\geq 0} (k^2-k)\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k\geq 2} k(k-1)\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\lambda^2\sum_{h\geq 2} \frac{\lambda^{k-2}}{(k-2)!} \\
&= \lambda^2.
\end{aligned}$$

The variance can now be computed as

$$\sigma_X^2 = \mathbb{E}X^2 - \mathbb{E}X + \mathbb{E}X - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

2 Expectation in terms of Probability

To explain their intimate connection, we now express the expectation of a $\mathbb{N}-$ valued random variable in terms of just the probabilities.

Theorem 1 *For a non-negative integer valued random variable X,*

$$\mathbb{E}[X] = \sum_{n\geq 1} P(X \geq n).$$

Proof:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{j\in\mathbb{N}} jP(X=j) \\
&= \sum_{j\in\mathbb{N}}\sum_{n=1}^{\infty} \mathbb{1}_{\{n\leq j\}}P(X=j) \\
&= \sum_{n=1}^{\infty}\sum_{j\in\mathbb{N}} \mathbb{1}_{\{n\leq j\}}P(X=j) \\
&= \sum_{n=1}^{\infty}\sum_{j\geq n} P(X=j) \\
&= \sum_{n=1}^{\infty} P(X\geq n).
\end{aligned}$$

■

To show the convenience of this property, let us consider the example of a popular random variable in the next subsection.

2.1 Geometric Random Variable- Geometric(p)

This random variable takes values in the countable state space $\mathbb{N}$, and the only parameter governing its realization is $p \in [0,1]$. In relation to tossing a coin, a geometric random variable X captures the first occurrence of HEADs. The probability of HEAD occurring on the k^{th} toss and not before it is

$$P(X = k) = (1-p)^{k-1}p,\ k \geq 1.$$

Example 1 *Compute* $\mathbb{E}[X]$ *and* $\mathbb{E}[X^2]$ *if* $X \sim$ *Geometric*(p).

Solution: Let us find $P(X \geq k)$ first.

$$\begin{aligned} P(X \geq k) &= \sum_{j=k}^{\infty}(1-p)^{j-1}p \\ &= p(1-p)^{k-1}\sum_{j=0}^{\infty}(1-p)^j \\ &= (1-p)^{k-1}. \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k\geq 1} P(X \geq k) \\ &= \sum_{k\geq 1}(1-p)^{k-1} \\ &= \frac{1}{p}. \end{aligned}$$

3 Markov's Inequality

To further underline that expectation contains significant information about the distribution, we state the **Markov's inequality**.

Theorem 2 *For a non-negative valued random variable* X*,*

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{1}$$

Proof: While the proof can be done by expanding the summation of expectation, we resort to a more elegant method using indicator functions.

$$\begin{aligned} X &= X\,\mathbb{1}_{\{X<a\}} + X\,\mathbb{1}_{\{X\geq a\}} \\ &\geq X\,\mathbb{1}_{\{X\geq a\}} \\ &\geq a\,\mathbb{1}_{\{X\geq a\}}. \end{aligned}$$

Taking expectation of both sides, we get

$$aP(X \geq a) \leq \mathbb{E}[X].$$

■

4 More On Independence

Theorem 3 *Consider the random vector* $(X_1, X_2) \in E_1 \times E_2$*. Then* $X_1.X_2$ *is a random variable.*

Proof: Let $Y = X_1.X_2$, we will show that Y is a random variable. Clearly Y is discrete. Furthermore,

$$\begin{aligned}\{Y = y\} &= \{X_1.X_2 = y\} \\ &= \bigcup_x \{(\omega_1, \omega_2) : X_1(\omega_1) = x, X_2(\omega_2) = \frac{y}{x}\}\end{aligned}$$

Since X_1 and X_2 are measurable maps, the above union is well defined and an element of $\mathcal{F}$. Hence Y is a random variable.

■

Theorem 4 *Consider the random variable* $(X_1, X_2) \in E_1 \times E_2$*, let* $g_1 : E_1 \to \bar{\mathbb{R}}$ *and* $g_2 : E_2 \to \bar{\mathbb{R}}$ *be two functions. Then* $g_1(X_1)g_2(X_2)$ *is a random variable.*

Proof: Since we know $g_1(X_1)$ and $g_2(X_2)$ are random variables, applying the previous theorem shows that $g_1(X_1).g_2(X_2)$ is indeed a random variable. ■

Let us now come back to our discussion of independence. The independence of random variables is preserved by individual functional transformations.

Theorem 5 *Let* X_1 *and* X_2 *be independent random variables in* E_1 *and* E_2 *respectively. Consider two functions* $g_1 : E_1 \to \mathbb{R}$ *and* $g_2 : E_2 \to \mathbb{R}$*. The random variables* $g_1(X_1)$ *and* $g_2(X_2)$ *are independent.*

Solution: The RVs $g_1(X_1)$ and $g_2(X_2)$ are discrete. Considering joint events,

$$\{g_1(X_1) = y_1, g_2(X_2) = y_2\} = \bigcup_{(x_1,x_2):g_1(x_1)=y_1,g_2(x_2)=y_2} \{X_1 = x_1, X_2 = x_2\},$$

where the events inside the union are disjoint. The probability of the event on the LHS is,

$$\begin{aligned}P(g_1(X_1) = y_1, g_2(X_2) = y_2) &= \sum_{\substack{(x_1,x_2):\\ g_1(x_1) = y_1,\\ g_2(x_2) = y_2}} P(X_1 = x_1, X_2 = x_2) \\ &= \sum_{\substack{(x_1,x_2):\\ g_1(x_1) = y_1,\\ g_2(x_2) = y_2}} P(X_1 = x_1)P(X_2 = x_2) \\ &= \sum_{x_1:g_1(x_1)=y_1} P(X_1 = x_1) \sum_{x_2:g_2(x_2)=y_2} P(X_2 = x_2) \\ &= P(g_1(X_1) = y_1)P(g_2(X_2) = y_2).\end{aligned}$$

Thus, by the definition of independence of two random variables, $g_1(X_1)$ and $g_2(X_2)$ are independent. ■

Note: There is some room for an argument that the correct proof should show

$$P(g_1(X_1) = y_1, g_2(X_2) = y_2) = P_1(g_1(X_1) = y_1)P_2(g_2(X_2) = y_2),$$

where $P_1(\cdot)$ and $P_2(\cdot)$ are the corresponding measures. Indeed we have shown this desired result, but the notation that we used in the theorem is an universally accepted one. In particular, we may use the same notation $P(\cdot)$ for different probability measures, but depending on the argument, the associated measure will be understood. For example, $P(X_1 = x_1)$ is the measure induced by X_1 on E_1, where as $P(X_2 = x_2)$ is a possibly different measure, induced by the random variable X_2.

Observe that the transformations were applied in a separable fashion on each variable. Let us now state a simple, yet powerful law for the distribution of expectation over product of independent random variables.

Proposition 3 *The independent random variables X and Y*

$$\mathbb{E}[g_1(X)g_2(Y)] = \mathbb{E}[g_1(X)][\mathbb{E}g_2(Y)]$$

whenever these expectations are well defined.

Proof:

$$\begin{aligned}
\mathbb{E}g_1[X]g_2[Y] &= \sum_{x,y} g_1(x)g_2(y)P(X = x, Y = y) && (2)\\
&= \sum_{x,y} g_1(x)g_2(y)P(X = x)P(Y = y) && (3)\\
&= \sum_{x \in E_1} g_1(x)P(X = x) \sum_{y \in E_2} g_2(y)P(Y = y) && (4)\\
&= \mathbb{E}[g_1(X)]\mathbb{E}[g_2(Y)]. && (5)
\end{aligned}$$

Example 2 *A fair die is thrown k times. Let X_i be the outcome of the i^{th} toss. Consider the probability association,*

$$P(X_1 = x_1, \cdots, X_k = x_k) = \frac{1}{6^k}.$$

Show that the random variables $X_1, \cdots, X_k$ are pairwise independent, i.e. every pair of X_i, X_j with $i \neq j$ are independent.

Solution: Clearly each X_i takes values in $\{1, \cdots, 6\}$. Let us consider the event $\{X_i = x_i, X_j = x_j\}$.

$$\begin{aligned}
P(X_i = x_i, X_j = x_j) &= \sum_{x_1, \cdots, x_m \setminus \{x_i, x_j\}} P(x_1, \cdots, x_k)\\
&= \sum_{x_1, \cdots, x_m \setminus \{x_i, x_j\}} \frac{1}{6^k}\\
&= \frac{1}{6^k} 6^{k-2}\\
&= \frac{1}{36}.
\end{aligned}$$

Thus, $P(X_i = x_i, X_j = x_j) = P(X_i = x_i)P(X_j = x_j)$. See footnote[1] below.

[1] we used $x_1, \cdots, x_k \setminus \{x_i, x_j\}$ to take the sum of over all subscripts except i and j

Example 3 *Consider the above example, and let $S_k = \sum_{i=1}^{k} X_i$. Compute the mean and variance of S_k.*

Solution: Let the mean be denoted as μ_k.

$$\mu_k = \mathbb{E}[S_k] = \sum_{i=1}^{k} \mathbb{E}[X_i] = \frac{7}{2}k.$$

Let us compute the variance $\sigma_{S_k}^2 = \mathbb{E}[(S_k - \mu_k)^2]$.

$$\mathbb{E}[(S_k - \mu_k)^2] = \mathbb{E}[\left(\sum_{i=1}^{k}(X_i - \mathbb{E}[X_i])\right)^2] \tag{6}$$

$$= \mathbb{E}[\sum_{i=1}^{k}(X_i - \mathbb{E}[X_i])^2 + \sum_{i,j:i\neq j}(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \tag{7}$$

$$= \sum_{i=1}^{k}\mathbb{E}(X_i - \mathbb{E}[X_i])^2 + \sum_{i,j:i\neq j}\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \tag{8}$$

$$= \sum_{i=1}^{k}\sigma_{X_i}^2 + \sum_{i,j:i\neq j}\mathbb{E}[X_i - \mathbb{E}[X_i]]\mathbb{E}[X_j - \mathbb{E}[X_j]] \tag{9}$$

$$= \sum_{i=1}^{k}\sigma_{X_i}^2 \tag{10}$$

Since $\sigma_{X_i}^2 = \frac{70}{24}, \forall i$, we have $\sigma_{S_k}^2 = \frac{70k}{24}$. ■

In fact, our computations in the last example show that we can generalize this result to arbitrary random variables which are pair-wise independent.

Theorem 6 *Let $X_1, \cdots, X_k$ be a collection of pairwise independent random variables. Then*

$$\sigma_{X_1+\cdots+X_k}^2 = \sum_{i=1}^{k}\sigma_{X_i}^2.$$

We have already defined independent sequence of random variables. Among this, a a particular class is the most appealing.

4.1 IID Random Variables

Definition 1 *A sequence $X_n, n \geq 1$ of RVs is said to be Independent and Identically Distributed (IID) if*

1. *$X_n, n \geq 1$ is a independent sequence, each X_n taking values in the same set E.*
2. *$P(X_i = x) = P(X_j = x), \forall i, j$ and $\forall x \in E$.*