10. Lectures 10 and 11.

All random variables below are assume discrete. For a random variable X we denote its range by $R(X)$.

10.1. **Some basic results about mean and variance.**

Theorem 10.1. *Let X and Y be random variables defined on the same sample space. Then $E(X+Y) = EX + EY$.*

Proof. We will first give a proof directly from definition and then explain how it could be shortened it using suitable more general theorems.

By definition of expectation

$$
\begin{aligned}
(10.1)\quad E(X+Y) &= \sum_{z\in R(X+Y)} z\cdot P(X+Y=z) \\
&= \sum_{z\in R(X+Y)} \sum_{z=x+y,\, x\in R(X),\, y\in R(Y)} z\cdot P(X=x, Y=y) \\
&= \sum_{z\in R(X+Y)} \sum_{z=x+y,\, x\in R(X),\, y\in R(Y)} (x+y)P(X=x, Y=y)
\end{aligned}
$$

(Here $P(X=x, Y=y)$ is the probability of the event ($X=x$ and $Y=y$)). In the last two expressions the inner summation (for a fixed z) is over all pairs (x,y) with $x\in R(X), y\in R(Y)$ such that $x+y=z$.

Now observe that each pair (x,y) with $x\in R(X), y\in R(Y)$ contributes exactly one term in the last expression of (10.1), and the contribution is equal to $(x+y)P(X=x,Y=y)$. Therefore, the last expression can be rewritten as $\sum_{x\in R(X), y\in R(Y)} (x+y)P(X=x,Y=y)$ which is equal to $\sum_{x\in R(X), y\in R(Y)} xP(X=x,Y=y) + \sum_{x\in R(X), y\in R(Y)} yP(X=x,Y=y)$. Now we evaluate each of these two summations separately.

First note that

$$
\begin{aligned}
&\sum_{x\in R(X), y\in R(Y)} xP(X=x, Y=y) \\
&= \sum_{x\in R(X)} \Big(\sum_{y\in R(Y)} xP(X=x,Y=y)\Big) = \sum_{x\in R(X)} x \sum_{y\in R(Y)} P(X=x,Y=y).
\end{aligned}
$$

If we fix $x\in R(X)$, the events $(X=x \,\&\, Y=y)$ for different y are disjoint, and their union is clearly the event $X=x$. Therefore,

$$
\sum_{x\in R(X)} x \sum_{y\in R(Y)} P(X=x,Y=y) = \sum_{x\in R(X)} xP(X=x) = EX.
$$

By the same argument $\sum\limits_{x\in R(X),y\in R(Y)} yP(X = x, Y = y) = EY$, and putting everything together, we conclude that $E(X+Y) = EX + EY$. □

Note that the first part of the above proof was establishing the formula

$$E(X+Y) = \sum_{x\in R(X),y\in R(Y)} (x+y)P(X = x, Y = y). \qquad (**)$$

Instead of proving this directly, we could refer to a more general theorem (see [BT, p.94])

Theorem 10.2. *If X and Y are random variables on the same sample space, then for any function $g : \mathbb{R}^2 \to \mathbb{R}$ we have*

$$E(g(X,Y)) = \sum_{x\in R(X),y\in R(Y)} g(x,y)P(X = x, Y = y).$$

(To justify (**) we could simply apply Theorem 10.2 to $g(x,y) = x+y$).

Theorem 10.2 is a natural generalization of the corresponding theorem in the case of one random variable:

Theorem 10.3. *For any random variable X and any function $g : \mathbb{R} \to \mathbb{R}$ we have*

$$E(g(X)) = \sum_{x\in R(X)} g(x)P(X = x).$$

Theorem 10.3 is proved in [BT, p.84], and Theorem 10.2 can be proved by a similar method. Note that [BT] uses abbreviated notation for PMF of a random variable ($p_X(x)$ instead of $P(X = x)$) and for joint PMF of two random variables ($p_{X,Y}(x,y)$ instead of $P(X = x, Y = y)$).

Theorem 10.4 (Linearity of expectation)**.** *If X is a random variable and $a, b \in \mathbb{R}$ are constants, then $E(aX + b) = a \cdot EX + b$*

Proof. Exercise. □

Theorem 10.5. *Let X be a random variable. Then*

$$Var(X) = E(X^2) - (EX)^2.$$

Proof. By definition $Var(X) = E((X-EX)^2) = E(X^2 - 2EX\cdot X + (EX)^2)$. By Theorem 10.1,

$$E(X^2 - 2EX \cdot X + (EX)^2) = E(X^2) + E(-2EX \cdot X + (EX)^2).$$

Since $-2EX$ and $(EX)^2$ are constants, applying Theorem 10.4 with $a = -2EX$ and $b = (EX)^2$, we get $E(-2EX \cdot X + (EX)^2) = (-2EX) \cdot EX + (EX)^2 = -2(EX)^2 + (EX)^2 = -(EX)^2$.

Putting everything together, we conclude that $Var(X) = E(X^2) + (-(EX)^2) = E(X^2) - (EX)^2$. □

Theorem 10.6 (Mean of a geometric random variable)**.** *Let $p \in (0,1)$ and X a geometric random variable with parameter p. Then $EX = \frac{1}{p}$.*

Proof. By definition $R(X) = \mathbb{N}$ and $P(X = k) = p(1-p)^{k-1}$. Hence

$$EX = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

Starting with the equality

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \text{ for } x \in (-1,1)$$

(which is just the formula for the sum of a geometric series) and taking derivatives of both sides (differentiating RHS term-by-term), we get

$$\sum_{k=1}^{\infty} kx^{k-1} = ((1-x)^{-1})' = (-1)^2(1-x)^{-2} = \frac{1}{(1-x)^2} \text{ for } x \in (-1,1). \tag{$***$}$$

(A general theorem asserts that if $\sum_{n=0}^{\infty} a_n(x-c)^n$ is a power series centered at x, R is its radius of convergence and $f(x) = \sum_{n=0}^{\infty} a_n(x-c)^n$ which is a function defined at least on of the open interval $(c-R, c+R)$, then $f'(x) = \sum_{n=1}^{\infty} na_n(x-c)^{n-1}$ for all $x \in (c-R, c+R)$).

Now setting $x = 1-p$ in (***) and multiplying both sides by p, we get $\sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}$, which completes the proof. □

Theorem 10.7 (Mean and variance of a Poisson random variable)**.** *Let $\lambda > 0$ be a real number and X a Poisson random variable with parameter p. Then $EX = Var(X) = \lambda$.*

Proof. By definition of Poisson random variable $R(X) = \mathbb{Z}_{\geq 0}$ (non-negative integers) and $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for each $k \in \mathbb{Z}_{\geq 0}$. Note that this is a legitimate PMF since $e^{-\lambda}\frac{\lambda^k}{k!} \geq 0$ and

$$\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

We start by computing the mean:

$$
\begin{aligned}
EX &= \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \tfrac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \tfrac{\lambda^k}{k!} \qquad k = 0 \text{ term vanishes} \\
&= \sum_{k=1}^{\infty} e^{-\lambda} \tfrac{\lambda \cdot \lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \tfrac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \tfrac{\lambda^j}{j!} \qquad \text{set } j = k-1 \\
&= \lambda e^{-\lambda} e^{\lambda} = \lambda.
\end{aligned}
$$

For the variance we will use the formula $Var(X) = E(X^2) - (EX)^2$. Applying Theorem 10.3 with $g(x) = x^2$, we get

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{(k-1)!}$$

Writing k as $k - 1 + 1$, we obtain that the last expression is equal to

$$\sum_{k=1}^{\infty} (k-1) e^{-\lambda} \frac{\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} = \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-2)!} + \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!}$$

The second summation in the last expression is simply EX (by an earlier computation) which we found to be equal to λ, and the first summation is equal to λ^2 by a similar argument (factor out λ^2 and make a change of variable $j = k - 2$).

Thus we showed that $E(X^2) = \lambda^2 + \lambda$, so $Var(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda$. □

10.2. **Independent Random Variables.**

Definition. Random variables X and Y (defined on the same sample space) are called independent if for every $x \in R(X)$ and $y \in R(Y)$ the events $X = x$ and $Y = y$ are independent. In other words, X and Y are independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y) \text{ for every } x \in R(X), y \in R(Y).$$

Example 1. *Suppose we toss a fair coin twice. Define random variables X and Y by*

$$X = \begin{cases} 1 & \textit{if heads on first toss} \\ 0 & \textit{if tails on first toss} \end{cases} \qquad Y = \begin{cases} 1 & \textit{if heads on second toss} \\ 0 & \textit{if tails on second toss} \end{cases}$$

Then X and Y are independent. Indeed, by definition, we need to check four equalities:

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1)$$
$$P(X = 1, Y = 0) = P(X = 1)P(Y = 0)$$
$$P(X = 0, Y = 1) = P(X = 0)P(Y = 1)$$
$$P(X = 0, Y = 0) = P(X = 0)P(Y = 0)$$

All these equalities are indeed true since by definition of a fair coin LHS in each of the four cases is equal to $\frac{1}{4}$ and RHS is equal to $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ as well.

On the other hand, if we set $Z = X + Y$, then X and Z are not independent. For instance, $P(X = 0, Z = 2) = 0$ since $Z = 2$ forces $X = Y = 1$ (so the event $X = 0 \,\&\, Z = 2$ cannot happen) while $P(X = 0)P(Z = 2) = P(X = 0)P(X = Y = 1) = \frac{1}{2} \cdot \frac{1}{4} \neq 0$.

Theorem 10.8. *Let X and Y be* **independent** *random variables defined on the same sample space. Then $E(XY) = EX \cdot EY$.*

Proof. Exercise (proof is similar to that of Theorem 10.1). □

Theorem 10.9 (Variance of the sum of independent random variables)**.** *Let X and Y be* **independent** *random variables defined on the same sample space. Then $Var(X + Y) = Var(X) + Var(Y)$.*

Proof. Using Theorems 10.5, 10.1 and 10.4, we have

$$Var(X+Y) = E((X+Y)^2)-(E(X+Y))^2 = E(X^2+2XY+Y^2)-(EX+EY)^2$$
$$= E(X^2) + 2E(XY) + E(Y^2) - ((EX)^2 + 2EX \cdot EY + (EY)^2)$$
$$= (E(X^2)-(EX)^2)+(E(Y^2)-(EY)^2)+2(E(XY)-EX\cdot EY) = Var(X)+Var(Y)+0$$

(where the last term is 0 by Theorem 10.8). □

Independence of more than two random variables. A finite collection of random variables $X_1, \ldots, X_n$ is called independent if for every $x_1 \in R(X_1), x_2 \in R(X_2), \ldots, x_n \in R(X_n)$, the events

$$X_1 = x_1, \quad X_2 = x_2, \quad \ldots, \quad X_n = x_n$$

are independent.

For instance, three random variables X, Y, Z are independent if for every $x \in R(X), y \in R(Y)$ and $z \in R(Z)$ we have

(1) $P(X = x, Y = y) = P(X = x)P(Y = y)$
(2) $P(X = x, Z = z) = P(X = x)P(Z = z)$
(3) $P(Y = y, Z = z) = P(Y = y)P(Z = z)$
(4) $P(X = x, Y = y, Z = z) = P(X = x)P(Y = y)P(Z = z)$.