Module 10.1: Independent Random Variables

MCIT Online - CIT592 - Professor Val Tannen

LECTURE NOTES

Penn Engineering

Independent random variables

Random variables X and Y defined on the same probability space are **independent**, written $X \perp Y$, when

$$\forall x \in \mathrm{Val}(X) \ \ \forall y \in \mathrm{Val}(Y) \quad (X = x) \perp (Y = y)$$

Example. Green-purple fair dice are rolled, G returns the number shown by the green die, P by the purple die. Then $G \perp P$, because for any $g, p \in [1..6]$:

$$\Pr[(G = g) \cap (P = p)] = 1/36 = (1/6)(1/6) = \Pr[G = g] \cdot \Pr[P = p].$$

When we have three or more r.v.'s, we define again **pairwise** and **mutual independence** analogously.

Independence of indicators and events

Proposition. Let A, B be two events in the probability space $(\Omega, \Pr)$. Then $I_A \perp I_B$ iff $A \perp B$

Proof. First we show $I_A \perp I_B \Rightarrow A \perp B$.

Assume $I_A \perp I_B$. By definition, $(I_A = 1) \perp (I_B = 1)$. Hence $A \perp B$.

Next, we show $A \perp B \Rightarrow I_A \perp I_B$.

If $A \perp B$ then by **Ind (iv)** we have $\overline{A} \perp \overline{B}$, $\overline{A} \perp B$, and $A \perp \overline{B}$. These cover all the cases needed for the definition of $I_A \perp I_B$.

This proposition can be extended straightforwardly to **mutual independence** of an arbitrary number of r.v.'s (we omit the statement and proof).

When we considered multiple IID Bernoulli trials "performed independently" we assumed that the corresponding Bernoulli random variables are **mutually independent**.

Quiz

A fair coin is tossed twice. Let X_H be random a variable that returns the number of heads observed. Clearly $\mathrm{Val}(X_H) = \{0, 1, 2\}$. Which one is the **bigger** probability?

(A) $\Pr[X_H = 0]$

(B) $\Pr[X_H = 1]$

(C) $\Pr[X_H = 2]$

Answer

(A) $\Pr[X_H = 0]$

Incorrect. There is only one outcome in which there are no heads.

(B) $\Pr[X_H = 1]$

Correct. There are **two** outcomes in which there is exactly one head.

(C) $\Pr[X_H = 2]$

Incorrect. There is only one outcome in which there are two heads.

More information

The probability space is uniform with 4 outcomes: $\{\texttt{HH},\ \texttt{HT},\ \texttt{TH},\ \texttt{TT}\}$.

The events for the distribution and their probabilities are

$(X_H = 0) = \{\texttt{TT}\}$ with $\Pr[X_H = 0] = 1/4$

$(X_H = 1) = \{\texttt{HT},\ \texttt{TH}\}$ with $\Pr[X_H = 1] = 1/2$

$(X_H = 2) = \{\texttt{HH}\}$ with $\Pr[X_H = 2] = 1/4$

Two r.v.'s that are not independent

Problem. A fair coin is flipped twice. Let X_H and X_T be random variables that return, respectively, the number of heads and tails observed. Are X_H and X_T independent?

Answer. Intuitively, the r.v.'s are **not** independent. For example, $X_H = 1$ **forces** $X_T = 1$. Let's verify this in detail.

The probability space on which the two r.v.'s are defined is uniform and has the set of outcomes $\{\mathtt{HH}, \mathtt{HT}, \mathtt{TH}, \mathtt{TT}\}$ each with probability $1/4$.

Then, $\Pr[X_T = 1 \,|\, X_H = 1] \;=\; \Pr[(X_T = 1) \cap (X_H = 1)] \,/\, \Pr[X_H = 1]$.

Note that $(X_T = 1) \cap (X_H = 1) = \{\mathtt{HT}, \mathtt{TH}\} = (X_H = 1) = (X_T = 1)$.

Thus, $\Pr[X_T = 1 \,|\, X_H = 1] \;=\; (1/2)/(1/2) = 1$. Also, $\Pr[X_T = 1] = 1/2$.

It follows that $(X_T = 1) \not\perp (X_H = 1)$ and therefore $X_H \not\perp X_T$.

ACTIVITY : Constant r.v.'s are independent of any r.v.

In this activity we prove the following

Proposition, For any probability space $(\Omega, \Pr)$, any random variable $X : \Omega \to \mathbb{R}$, and any $c \in \mathbb{R}$, the constant r.v. $C : \Omega \to \mathbb{R}$ defined by $\forall w \in \Omega \quad C(w) = c$ is independent of X, $C \perp X$.

Proof. Since $\mathrm{Val}(C) = \{c\}$ all we have to prove is

$$\forall x \in \mathrm{Val}(X) \quad (X = x) \perp (C = c)$$

This follows from the observation that $(C = c) = \Omega$ together with property **Ind (ii)** in lecture segment "Independence" which states that Ω is independent of any event.

Module 10.2: Variance

MCIT Online - CIT592 - Professor Val Tannen

Lecture Notes

Measuring deviation from mean

If X is an r.v. its **expectation**, $\mathrm{E}[X]$, also called its **mean**, is often denoted by $\mu = \mathrm{E}[X]$.

The **deviation** of X from its mean, $X - \mu$ is also an r.v.

However, if we take $\mathrm{E}[X - \mu]$ we get, by the linearity of expectation $\mathrm{E}[X] - \mathrm{E}[\mu] = \mu - \mu = 0$!

This is not very informative. We do not want the positive and the negative deviations from the mean to cancel each other out.

We could take the expectation of the absolute value of $X - \mu$.

However, working with absolute values is mathematically messy.

As a result, statisticians decided that squaring $X - \mu$ is more useful!

Variance

The **variance** of a random variable X is defined as

$$\mathrm{Var}[X] \;=\; \mathrm{E}[(X-\mu)^2] \qquad (\text{where } \mu = \mathrm{E}[X])$$

The **standard deviation** of a random variable X is

$$\sigma[X] \;=\; \sqrt{\mathrm{Var}[X]}$$

Why square root? By undoing the squaring in the variance we obtain the same units of measurement as those used for the values of the random variable.

Notation. When the random variable is understood , its mean is often denoted by μ, its standard deviation by σ, and its variance by σ^2.

An alternative formula for variance

Proposition. Let X be an r.v. defined on $(\Omega, \Pr)$.

$$\mathrm{Var}[X] \;=\; \mathrm{E}[X^2] - \mu^2 \qquad\qquad (\text{where } \forall w \in \Omega \;\; X^2(w) = (X(w))^2\ )$$

Proof. Using linearity of expectation and the fact that μ is a constant:

$$\begin{aligned}
\mathrm{E}[(X-\mu)^2] &= \mathrm{E}[X^2 - 2\mu X + \mu^2] \\
&= \mathrm{E}[X^2] - 2\mu \mathrm{E}[X] + \mu^2 \\
&= \mathrm{E}[X^2] - 2\mu^2 + \mu^2 \;=\; \mathrm{E}[X^2] - \mu^2
\end{aligned}$$

Activity : Variance of a fair die

In the lecture segment "Random variables" we introduced the random variable D that returns the number shown by a fair die. This is a uniform r.v. with $\text{Val}(D) = [1..6]$. Later, in the lecture segment "Expectation" we calculated $\text{E}[D] = 3.5$. In this activity we will calculate $\text{Var}[D]$, the **variance** of the number shown by a fair die.

Question. What is the distribution of the r.v. D^2?

In the video, there is a box here for learners to put in an answer to the question above. As you read these notes, try it yourself using pen and paper!

ACTIVITY : Variance of a fair die (continued)

Answer. Like D, the r.v. D^2 is **uniform**. We have $\mathrm{Val}(D^2) = \{1, 4, 9, 16, 25, 36\}$ and the probability is $1/6$ for each of these.

This is a particular case of a more general fact: If X is an r.v. that takes positive values then the distribution of X^2 is very similar to that of X, with the same probabilities, but the values taken are squared.

Activity : Variance of a fair die (continued)

Now to the calculation of Var[D]. Using the alternative formula for variance we have

$$\begin{aligned}
\mathrm{Var}[D] &= \mathrm{E}[D^2] - \mathrm{E}[D]^2 \\
&= \frac{1}{6}(1+4+9+16+25+36) - (3.5)^2 \\
&= \frac{35}{12} \simeq 2.92
\end{aligned}$$

Module 10.3: Linearity of Variance?

MCIT Online - CIT592 - Professor Val Tannen

Lecture Notes

Variance of Bernoulli r.v.'s

Problem. Let X be a Bernoulli r.v. with parameter $\Pr[X = 1] = p$. Calculate $\text{Var}[X]$.

Answer. We have already calculated the expectation of a Bernoulli r.v. as $\mu = \text{E}[X] = p$.

To compute the variance we will need the expectation of X^2. For any outcome w we have:

. $\quad X^2(w) = 1$ iff $X(w) = 1$ and $X^2(w) = 0$ iff $X(w) = 0$.

Therefore, X^2 is also Bernoulli, with the same distribution. In fact, X^2 equals X!

We conclude that $\text{Var}[X] = \text{E}[X^2] - \mu^2 = p - p^2 = p(1-p)$.

Quiz

Let X be a Bernoulli r.v. with parameter $1/3$ and let $Y = 1 - X$. Then

(A) X and Y have the **same** expectation and the **same** variance.

(B) X and Y have the **same** expectation and **different** variances.

(C) X and Y have **different** expectations and the **same** variance.

Answer

(A) X and Y have the **same** expectation and the **same** variance.
Incorrect. Observe that Y is also Bernoulli but with parameter $1 - (1/3) = 2/3$. Hence $\mathrm{E}[X] = 1/3$ but $\mathrm{E}[Y] = 2/3$.

(B) X and Y have the **same** expectation and **different** variances.
Incorrect. Expectations are different, see (A).

(C) X and Y have **different** expectations and the **same** variance.
Correct. Observe that Y is also Bernoulli but with parameter $1 - (1/3) = 2/3$. Therefore
$\mathrm{Var}[X] = (1/3)(1 - (1/3)) = (2/3)(1 - (2/3)) = \mathrm{Var}[Y]$

Variance of the sum of two fair dice

Problem. Recall S the r.v. that returns the sum of numbers shown by two fair dice rolled together. Calculate $\mathrm{Var}[S]$.

Answer. In an earlier segment we have calculated $\mathrm{E}[S] = 7$. It remains to calculate $\mathrm{E}[S^2]$.

Like S, S^2 takes 11 values. Namely, the squares of the 11 values of S with the same probabilities. We show only some of the 11 terms in the variance sum:

$$\mathrm{E}[S^2] \;=\; 4 \cdot (1/36) + \cdots + 25 \cdot (4/36) + \cdots + 49 \cdot (6/36) + \cdots + 144 \cdot (1/36)$$

Can we avoid this calculation...?

From previous segments, we know that $S = G + P$ and we calculated $Var[G] = \mathrm{Var}[P] = \mathrm{Var}[D] = 35/12$. Is there **linearity of variance**?

Linearity of variance?

Proposition. $\text{Var}[c\,X] \;=\; c^2\,\text{Var}[X]$

Proof. $\text{Var}[c\,X] \;=\; \text{E}[(cX)^2] - (\text{E}[cX])^2 \;=\; \text{E}[c^2\,X^2] - (c\,\text{E}[X])^2$

$$= \; c^2\,\text{E}[X^2] - c^2(\text{E}[X])^2 \;=\; c^2(\text{E}[X^2] - (\text{E}[X])^2) \;=\; c^2\,\text{Var}[X]$$

In general, variance does not distribute over sums. However:

Proposition. if $X \perp Y$ then $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$.

Proof. In the segment entitled "Correlated random variables" we define **product** of r.v.'s and show

1) $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ iff $\text{E}[XY] = \text{E}[X]\text{E}[Y]$
2) $X \perp Y \;\Rightarrow\; \text{E}[XY] = \text{E}[X]\text{E}[Y]$. The proposition follows.

When we roll two fair green-purple dice, we have $S = G + P$ and $G \perp P$.

Therefore, $\text{Var}[S] = Var[G] + \text{Var}[P] = (35/12) + (35/12) = 35/6$.

Module 10.4: Binomial Distribution

MCIT Online - CIT592 - Professor Val Tannen

Lecture Notes

Binomial random variables

An r.v. $B : \Omega \to \mathbb{R}$ is called **binomial** with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ when $\mathrm{Val}(B) = [0..n]$ and $\forall k \in [0..n]$ $\Pr[B = k] = \binom{n}{k} p^k (1-p)^{n-k}$.

How does such an r.v. arise? For example, perform n IID Bernoulli trials with probability of success p and let B be the r.v. that returns the number of successes observed. Clearly $\mathrm{Val}(B) = [0..n]$. Then:

We have seen before the probability space on which B is defined: the outcomes are the 2^n sequences of length n of $\mathtt{S}$'s (for "success") and $\mathtt{F}$'s (for "failure"). An outcome with k $\mathtt{S}$'s has probability $p^k(1-p)^{n-k}$.

There are $\binom{n}{k}$ outcomes with k $\mathtt{S}$'s. Therefore the probability of the event "k successes observed" is $\binom{n}{k} p^k (1-p)^{n-k}$.

The distribution of B is also called **binomial** with parameters n and p.

Penn Engineering

Examples of binomial

Example. We considered the r.v.'s X_H and X_T that return the number of heads, and respectively tails, shown when we flip a fair coin twice.

Both X_H and X_T are binomial r.v.'s with $n = 2$ and $p = 1/2$.

Example. We throw k balls into m bins. The r.v. that returns the number of balls that end up in Bin 1 is a binomial r.v. with $n = k$ and $p = 1/m$.

Example. We roll a fair die k times. The r.v. that returns the number of twos and threes shown is a binomial r.v. with $n = k$ and $p = 1/3$.

Expectation for binomial

Since expectation (and variance) are completely determined by the distribution it suffices to compute them for B be the (binomial) r.v. that returns the number of successes observed in n IID **Bernoulli** trials, each with probability of success p.

Let S_i be the event "the i'th trial resulted in $\mathtt{S}$". We have $\Pr[S_i] = p$.

Let J_i be the **indicator** random variable of the event S_i.
$\mathrm{E}[J_i] = \Pr[J_i = 1] = \Pr[S_i] = p$

Now, the binomial r.v. B can be **decomposed** into a sum of indicators:
$B = J_1 + \cdots + J_n$.

By linearity of expectation $\mathrm{E}[B] = \mathrm{E}[J_1] + \cdots + \mathrm{E}[J_n]$.

In conclusion $\mathrm{E}[B] = p + \cdots + p = np$.

Variance for binomial I

Yet again, let B be the (binomial) r.v. that returns the number of successes observed in n IID **Bernoulli** trials, each with probability of success p.

On the previous slide we decomposed $B = J_1 + \cdots + J_n$ where J_i is the **indicator** random variable of the event $S_i =$"the i'th trial resulted in S".

In general, variance does **not** distribute over sums. However, we have

Proposition. If r.v.'s $X_1, \ldots, X_n$ are **pairwise independent** then

$$\mathrm{Var}[X_1 + \cdots + X_n] \;=\; \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]$$

The **proof** (and a more general formulation) is in the segment "Correlated random variables".

Variance for binomial II

$B = J_1 + \cdots + J_n$ where (recall) J_i is the indicator random variable of the event $S_i =$"the i'th trial resulted in S".

The events S_i are mutually independent and therefore the indicator r.v.'s J_i are also mutually independent. We can apply the proposition:

$$\mathrm{Var}[J_1 + \cdots + J_n] \;=\; \mathrm{Var}[J_1] + \cdots + \mathrm{Var}[J_n]$$

We have calculated earlier the variance of Bernoulli r.v.'s with parameter p. Therefore, $\mathrm{Var}[J_i] = \Pr[J_i = 1](1 - \Pr[J_i = 1]) = p(1-p)$.

In conclusion, $\mathrm{Var}[B] = p(1-p) + \cdots + p(1-p) = np(1-p)$.