EECS 70: Lecture 27.

Joint and Conditional Distributions.

1. Recap of variance of a random variable
2. Joint distributions
3. Recap of indep. rand. variables: Variance of $B(n,p)$
4. Conditioning of Random Variables (revisit $G(p)$)

Recap

Variance

- Variance: $var[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Fact: $var[aX + b] = a^2 var[X]$
- Thm.: If X, Y are indep., $Var(X + Y) = Var(X) + Var(Y)$.
- $U[1, \dots, n] : Pr[X = m] = \frac{1}{n}, m = 1, \dots, n;$
  $E[X] = \frac{n+1}{2}; var(X) = \frac{n^2-1}{12}.;$
- $G(p) : Pr[X = n] = (1-p)^{n-1} p, n = 1, 2, \dots ;$
  $E[X] = \frac{1}{p}; var[X] = \frac{1-p}{p^2}.$
- $B(n, p) : Pr[X = m] = \binom{n}{m} p^m (1-p)^{n-m}, m = 0, \dots, n;$
  $E[X] = np; var(X) = \ = np(1-p).$

Joint distribution.

Two random variables, X and Y, in probability space: (Ω, P).

What is $\sum_x P[X = x]$? 1. What is $\sum_y P[Y = y]$? 1.

Let's think about: $P[X = x, Y = y]$.
What is $\sum_{x,y} P[X = x, Y = y]$?

Are the events "$X = x, \ Y = y$" disjoint?
Yes! Y and X are functions on Ω.
Do they cover the entire sample space?
Yes! X and Y are functions on Ω.

So, $\sum_{x,y} P[X = x, Y = y] = 1$.

Joint Distribution: $P[X = x, Y = y]$.
Marginal Distributions: $P[X = x]$ and $P[Y = y]$.
Important for inference.

Two random variables, same outcome space.

Experiment: pick a random person.
X = number of episodes of Games of Thrones they have seen.
Y = number of episodes of Westworld they have seen.

X	0	1	2	3	5	40	All
P	0.3	0.05	0.05	0.05	0.05	0.1	0.4

Is this a distribution?

Yes! All the probabilities are non-negative and add up to 1.

Y	0	1	5	10
P	0.3	0.1	0.1	0.5

Joint distribution: Example.

The **joint distribution** of X and Y is:

Y/X	0	1	2	3	5	40	All	
0	0.15	0	0	0	0	0.1	0.05	=0.3
1	0	0.05	0.05	0	0	0	0	=0.1
5	0	0	0	0.05	0.05	0	0	=0.1
10	0.15	0	0	0	0	0	0.35	=0.5
	=0.3	=0.05	=0.05	=0.05	=0.05	=0.1	=0.4	

Is this a valid distribution? Yes!
Notice that $P[X = a]$ and $P[Y = b]$ are (marginal) distributions!
But now we have more information!

For example, if I tell you someone watched 5 episodes of Westworld, they definitely didn't watch all the episodes of GoT.

Independent random variables.

Definition: Independence

The random variables X and Y are **independent** if and only if

$$P[Y = b \mid X = a] = P[Y = b], \text{ for all } a \text{ and } b.$$

Fact:

X, Y are independent if and only if

$$P[X = a, Y = b] = P[X = a]P[Y = b], \text{ for all } a \text{ and } b.$$

Don't need a huge table of probabilities like the previous slide.

Independence: examples.

Example 1

Roll two dices. X, Y = number of pips on the two dice. X, Y are independent.

Indeed: $P[X = a, Y = b] = 1/36, P[X = a] = P[Y = b] = 1/6$.

Example 2

Roll two dices. X = total number of pips, Y = number of pips on die 1 minus number on die 2. X and Y are not independent.

Indeed: $P[X = 12, Y = 1] = 0 \neq P[X = 12]P[Y = 1] > 0$.

Mean of product of independent RVs.

Theorem

Let X, Y be independent RVs. Then

$$E[XY] = E[X]E[Y].$$

Proof:

Recall that $E[g(X,Y)] = \sum_{x,y} g(x,y)P[X=x, Y=y]$. Hence,

$$\begin{aligned}
E[XY] &= \sum_{x,y} xyP[X=x, Y=y] = \sum_{x,y} xyP[X=x]P[Y=y], \text{ by ind.} \\
&= \sum_{x}\Big[\sum_{y} xyP[X=x]P[Y=y]\Big] \\
&= \sum_{x}\Big[xP[X=x]\Big(\sum_{y} yP[Y=y]\Big)\Big] \\
&= \sum_{x} xP[X=x]E[Y] = E[X]E[Y].
\end{aligned}$$

$\square$

Variance of sum of two independent random variables

Theorem:
If X and Y are independent, then

$$Var(X+Y) = Var(X) + Var(Y).$$

Proof:
Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E(X) = 0$ and $E(Y) = 0$.

Then, by independence,

$$E(XY) = E(X)E(Y) = 0.$$

Hence,

$$\begin{aligned} var(X+Y) &= E((X+Y)^2) = E(X^2 + 2XY + Y^2) \\ &= E(X^2) + 2E(XY) + E(Y^2) = E(X^2) + E(Y^2) \\ &= var(X) + var(Y). \end{aligned}$$

Examples.

(1) Assume that X, Y, Z are (pairwise) independent, with $E[X] = E[Y] = E[Z] = 0$ and $E[X^2] = E[Y^2] = E[Z^2] = 1$.

Then

$$\begin{aligned}
&E[(X + 2Y + 3Z)^2] \\
&\quad = E[X^2 + 4Y^2 + 9Z^2 + 4XY + 12YZ + 6XZ] \\
&\quad = 1 + 4 + 9 + 4 \times 0 + 12 \times 0 + 6 \times 0 \\
&\quad = 14.
\end{aligned}$$

(2) Let X, Y be independent and $U\{1, 2, \ldots, n\}$. Then

$$\begin{aligned}
E[(X - Y)^2] &= E[X^2 + Y^2 - 2XY] = 2E[X^2] - 2E[X]^2 \\
&= \frac{1 + 3n + 2n^2}{3} - \frac{(n+1)^2}{2}.
\end{aligned}$$

Variance: binomial.

$$\begin{aligned} E[X^2] &= \sum_{i=0}^{n} i^2 \binom{n}{i} p^i (1-p)^{n-i}. \\ &= \text{Really???!!##...} \end{aligned}$$

Too hard!

Ok.. fine.
Let's do something else.
Maybe not much easier...but there is a payoff.

Variance of Binomial Distribution.

Flip coin with heads probability p.
X- how many heads?

$$X_i = \begin{cases} 1 & \text{if } i\text{th flip is heads} \\ 0 & \text{otherwise} \end{cases}$$

$E(X_i^2) = 1^2 \times p + 0^2 \times (1-p) = p$.
$Var(X_i) = p - (E(X))^2 = p - p^2 = p(1-p)$.

$p = 0 \implies Var(X_i) = 0$
$p = 1 \implies Var(X_i) = 0$

$X = X_1 + X_2 + \ldots X_n$.

X_i and X_j are independent: $Pr[X_i = 1 | X_j = 1] = Pr[X_i = 1]$.

$$Var(X) = Var(X_1 + \cdots X_n) = np(1-p).$$

Conditioning of RVs

Recall conditioning on an event A

$$P[X = k \mid A] = \frac{P[(X = k) \cap A]}{P[A]}$$

Conditioning on another RV

$$P[X = k \mid Y = m] = \frac{P[X = k, Y = m]}{P[Y = m]} = p_{X|Y}(x \mid y)$$

$p_{X|Y}(x \mid y)$ is called the **conditional distribution or conditional probability mass function** (pmf) of X given Y

$$p_{X|Y}(x \mid y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

Conditional distributions

$X \mid Y$ is a RV:

$$\sum_x p_{X|Y}(x \mid y) = \sum_x \frac{p_{XY}(x,y)}{p_Y(y)} = 1$$

Multiplication or Product Rule:

$$p_{XY}(x,y) = p_X(x)p_{Y|X}(y \mid x) = p_Y(y)p_{X|Y}(x \mid y)$$

Total Probability Theorem: If A_1, A_2, $\ldots$, A_N partition Ω, and $P[A_i] > 0 \; \forall i$, then

$$p_X(x) = \sum_{i=1}^{N} P[A_i]P[X = x \mid A_i]$$

Nothing special about just two random variables, naturally extends to more.

Let's visit the mean and variance of the geometric distribution using conditional expectation.

Revisiting mean of geometric RV $X \sim G(p)$

X is **memoryless**

$$P[X = n + m \mid X > n] = P[X = m].$$

Thus $E[X \mid X > 1] = 1 + E[X]$.

Why? (Recall $E[g(X)] = \sum_l g(l)P[X = l]$)

$$\begin{aligned}
E[X \mid X > 1] &= \sum_{k=1}^{\infty} kP[X = k \mid X > 1] \\
&= \sum_{k=2}^{\infty} kP[X = k - 1] \qquad (\textit{memoryless}) \\
&= \sum_{l=1}^{\infty} (l + 1)P[X = l] \qquad (l = k - 1) \\
&= E[X + 1] = 1 + E[X]
\end{aligned}$$

Revisiting mean of geometric RV $X \sim G(p)$

X is **memoryless**

$$P[X = k + m \mid X > k] = P[X = m].$$

Thus $E[X \mid X > 1] = 1 + E[X]$.

We have $E[X] = P[X = 1]E[X \mid X = 1] + P[X > 1]E[X \mid X > 1]$.

$$\Rightarrow E[X] = p.1 + (1 - p)(E[X] + 1)$$
$$\Rightarrow E[X] = p + 1 - p + E[X] - pE[X]$$
$$\Rightarrow pE[X] = 1$$
$$\Rightarrow E[X] = \frac{1}{p}$$

Derive the variance for $X \sim G(p)$ by finding $E[X^2]$ using conditioning.

Summary of Conditional distribution

For Random Variables X and Y, $P[X = x \mid Y = k]$ is the **conditional distribution** of X given $Y = k$

$$P[X = x \mid Y = k] = \frac{P[X = x, Y = k]}{P[Y = k]}$$

Numerator: Joint distribution of (X, Y).

Denominator: Marginal distribution of Y.

(Aside: surprising result using conditioning of RVs):

Theorem: If $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$ are independent, then $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

"Sum of independent Poissons is Poisson."

Summary.

Joint and Conditional Distributions.

Joint distributions:

- Normalization: $\sum_{x,y} P[X=x, Y=y] = 1$.
- Marginalization: $\sum_{y} P[X=x, Y=y] = P[X=x]$.
- Independence: $P[X=x, Y=y] = P[X=x]P[Y=y]$ for all x, y. $E[XY] = E[X]E[Y]$.

Conditional distributions:

- Sum of independent Poissons is Poisson.
- Conditional expectation: useful for mean & variance calculations