Review

Conditional Probability

$$\mathbb{P}[A|B] \stackrel{\text{def}}{=} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

Example: **You go to the doctor for a routine checkup, in which she also administers a test for the condition C. The test is not 100% accurate; instead, the test shows positive (event TP) only with probability $.95$ if you actually do have the condition, and with probability $.05$ even if you don't:**

$$\mathbb{P}[TP|C] = 0.95 \text{ and } \mathbb{P}[TP|\overline{C}] = 0.05 .$$

In addition, the CDC estimates that 2%

of the people going for a routine checkup have the condition:

$$\mathbb{P}[C] = 0.02 \ .$$

Panic: The test shows positive!
What to do?

of the people going for a routine checkup have the condition:

$$\mathbb{P}[C] = 0.02 \ .$$

Panic: The test shows positive!

What to do?

You compute the probability that you actually have the condition:

$$\mathbb{P}[C|TP] = \frac{\mathbb{P}[TP|C]\cdot\mathbb{P}[C]}{\mathbb{P}[TP|C]\cdot\mathbb{P}[C]+\mathbb{P}[TP|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.95\times.02}{.95\times.02 + .05\times.98} \approx 28\% \ .$$

You feel a little relief but still want to know. You can go for a more accurate (and presumably more expensive) test, or take the same one over. The difference the second time around is that you now belong to a population having $\mathbb{P}[C] = .28$.

Let $T2P$ denote the event that the Second Test is Positive and compute:

$$\mathbb{P}[C|T2P] = \frac{\mathbb{P}[T2P|C]\cdot\mathbb{P}[C]}{\mathbb{P}[T2P|C]\cdot\mathbb{P}[C]+\mathbb{P}[T2P|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.95\times .28}{.95\times .28 + .05\times .72} \approx 88\% \ .$$

Now you are really worried. You may want to take the test a third time:

$$\mathbb{P}[C|T3P] = \frac{\mathbb{P}[T3P|C]\cdot\mathbb{P}[C]}{\mathbb{P}[T3P|C]\cdot\mathbb{P}[C]+\mathbb{P}[T3P|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.95\times .88}{.95\times .88 + .05\times .12} \approx 99\% \ .$$

Suppose however, the second test is negative. Then

$$\mathbb{P}[C|\overline{T2P}] = \frac{\mathbb{P}[\overline{T2P}|C]\cdot\mathbb{P}[C]}{\mathbb{P}[\overline{T2P}|C]\cdot\mathbb{P}[C]+\mathbb{P}[\overline{T2P}|\overline{C}]\cdot\mathbb{P}[\overline{C}]}$$

$$= \frac{.05\times .28}{.05\times .28 + .95\times .72} \approx 2\% \ .$$

You should be just as [un–]concerned about having C as you were before the first visit.

Independence

Definition: **Two events** $A, B \subseteq S$ **are independent, written** $A \perp\!\!\!\perp B$**, if**

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B] \, .$$

If $\mathbb{P}[B] \neq 0$ **we divide by it and find that** A, B **are independent if**

$$\mathbb{P}[A] = \mathbb{P}[A|B] \quad \Big(\textbf{ or } \mathbb{P}[B] = \mathbb{P}[B|A] \Big) \ :$$

i.e., if knowledge that B **has occurred does not alter the probability that** A **will occur and vice versa.**

Example: **Roll two distinguishable dice. The events** $[\text{die}_1 = 2]$ **and** $[\text{die}_2 = 5]$ **are**

independent. So are the events [die$_1$ = 2] and [sum of faces = 7].

$$\left\{\begin{matrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{matrix}\right\}$$

Two disjoint events of strictly positive probability are NOT independent.

Lemma: **If** $A \perp\!\!\!\perp B$ **then**

$$A \perp\!\!\!\perp \overline{B} \text{ and } \overline{A} \perp\!\!\!\perp B \text{ and } \overline{A} \perp\!\!\!\perp \overline{B}\ .$$

Proof:

$$\mathbb{P}[A \cap \overline{B}] = \mathbb{P}[A] - \mathbb{P}[A \cap B] = \mathbb{P}[A] - \mathbb{P}[A] \cdot \mathbb{P}[B]$$
$$= \mathbb{P}[A](1 - \mathbb{P}[B]) = \mathbb{P}[A] \cdot \mathbb{P}[\overline{B}] \,.$$

Use with $A \longrightarrow B$ and $A \longrightarrow \overline{A}$. QED

Definition: **The family $\{A_1, A_2, \ldots, A_n\} \subseteq \mathcal{P}[S]$ is independent if for every subfamily $\{A_{n_1}, \ldots, A_{n_k}\}$**

$$\mathbb{P}[\bigcap_{i=1}^{k} A_{n_i}] = \prod_{i=1}^{k} \mathbb{P}[A_{n_i}] \,.$$

This means that knowledge that any number of the A_i have occurred does not help reassessing the probability of any event A_j not in that number.

In our example of two dice the three events

$[\text{die}_1 = 2]$, $[\text{die}_2 = 5]$, $[\text{sum} = 7]$ are not independent, yet any pair among them is.

Example, Independent Trials: Suppose you have an experiment with finite space S and probability function

$$p(s) = \mathbb{P}[\{s\}] \qquad\qquad s \in S$$

— the "primitive" experiment — and you repeat it independently N times – independently means that the outcomes of any repetition are not influenced by the outcomes of any or all earlier or later repetitions. Then a reasonable state space for the new experiment that consists of the N repetitions of the "primitive" experiment is the N–fold cartesian Product

$$S^N \stackrel{\text{def}}{=} \{(s_1, s_2, \ldots, s_N) : s_i \in S\}$$

and the probability function on it is given by

$$p\big((s_1, s_2, \ldots, s_N)\big) \stackrel{\text{def}}{=} p(s_1) \cdot p(s_2) \cdots p(s_N) \; .$$

Example, Independent Bernoulli Trials: **A Bernoulli trial is an experiment with only two outcomes H and T such that $\mathbb{P}[\{H\}] = p$ and $\mathbb{P}[\{T\}] = 1 - p$. Repeat it N times. A typical outcome is**

$$\underbrace{HHT \cdots TTTH}_{N \text{ letters}} \; .$$

What is the probability of $[\#$ of $H = k]$?

by

$$p\big((s_1, s_2, \ldots, s_N)\big) \stackrel{\text{def}}{=} p(s_1) \cdot p(s_2) \cdots p(s_N) \; .$$

Example, Independent Bernoulli Trials: A Bernoulli trial is an experiment with only two outcomes H and T such that $\mathbb{P}[\{H\}] = p$ and $\mathbb{P}[\{T\}] = 1 - p$. Repeat it N times. A typical outcome is

$$\underbrace{HHT \cdots TTTH}_{N \text{ letters}} \; .$$

What is the probability of $[\# \text{ of } H = k]$? An outcome like

$$\underbrace{HHT \cdots TTTH}_{k \; H\text{s and } N-k \; T\text{s}} \qquad .$$

has probability $p^k(1-p)^{N-k}$. There are $\binom{N}{k}$

of them:

$$\mathbb{P}[\# \textbf{ of } H = k] = \binom{N}{k} p^k (1-p)^{N-k} \ .$$

This is the Binomial Distribution.
Keep repeating the Bernoulli experiment infinitely often. What is the probability of the event E that the outcome H appears first on the k^{th} repetition?

of them:

$$\mathbb{P}[\#\ \textbf{of}\ H = k] = \binom{N}{k} p^k (1-p)^{N-k}\ .$$

This is the Binomial Distribution.
Keep repeating the Bernoulli experiment infinitely often. What is the probability of the event E that the outcome H appears first on the $k^{\textbf{th}}$ repetition?

$$\mathbb{P}[E] = (1-p)^{k-1} p\ .$$

Bernoulli Trials

A Bernoulli trial is an experiment with only two outcomes H and T such that $\mathbb{P}[\{H\}] = p$ and $\mathbb{P}[\{T\}] = 1 - p$. Repeat it independently n times. A typical outcome is

$$\underbrace{HHT \cdots TTTH}_{n \text{ letters}} .$$

What is the probability of $[\#$ of $H = k]$?

Bernoulli Trials

A Bernoulli trial is an experiment with only two outcomes H and T such that $\mathbb{P}[\{H\}] = p$ and $\mathbb{P}[\{T\}] = 1 - p$. Repeat it independently n times. A typical outcome is

$$\underbrace{HHT\cdots TTTH}_{n \textbf{ letters}} \ .$$

What is the probability of $[\#$ of $H = k]$? An outcome like

$$\underbrace{HHT\cdots TTTH}_{k\ H\textbf{s and } n-k\ T\textbf{s}} \quad .$$

has probability $p^k(1-p)^{n-k}$. There are $\binom{n}{k}$

of them:

$$\mathbb{P}[\#\ \textbf{of}\ H = k] = \binom{n}{k} p^k (1-p)^{n-k}\ .$$

This is the Binomial Distribution.
Keep repeating the Bernoulli experiment infinitely often. What is the probability of the event E that the outcome H appears first on the $k^{\textbf{th}}$ repetition?

of them:

$$\mathbb{P}[\#\ \textbf{of}\ H = k] = \binom{n}{k} p^k (1-p)^{n-k}\ .$$

This is the Binomial Distribution.
Keep repeating the Bernoulli experiment infinitely often. What is the probability of the event E that the outcome H appears first on the $k^{\textbf{th}}$ repetition?

$$\mathbb{P}[E] = (1-p)^{k-1} p\ .$$

Random Variables

Definition: **Suppose as always that the state space S is finite or countable. A function $X : S \to \mathbb{R}$ is called a Random Variable. The Expectation of X is the number**

$$\mathbb{E}[X] \stackrel{\text{def}}{=} \sum_{s \in S} X(s) \cdot p(s) = \sum_{x \in \mathbb{R}} x \cdot \mathbb{P}[X = x] \,.$$

Frequently we write μ_X for $\mathbb{E}[X]$.

Proof of the equality:

$$S = \bigcup_{x \in \mathbb{R}} \{s \in S : X(s) = x\} = \bigcup_{x \in \mathbb{R}} [X = x] \,.$$

Hence

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x\in\mathbb{R}} \sum_{\{s:X(s)=x\}} X(s)p(s) \\ &= \sum_{x\in\mathbb{R}} \sum_{\{s:X(s)=x\}} xp(s) \\ &= \sum_{x\in\mathbb{R}} x \sum_{\{s:X(s)=x\}} p(s) \\ &= \sum_{x\in\mathbb{R}} x \cdot \mathbb{P}[X = x] \, .\end{aligned}$$

Theorem: The expectation is linear and positive: if X, Y are random variables and a, b constants then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

and

$$X \geq 0 \implies \mathbb{E}[X] \geq 0 \, .$$

Example: The Indicator Function of a set $A \subseteq S$ is defined by

$$\begin{cases} I_A(s) = 1 & \textbf{if } s \in A \, , \\ I_A(s) = 0 & \textbf{if } s \notin A \, . \end{cases}$$

Clearly

$$\mathbb{P}[A] = \mathbb{E}[I_A] \, .$$

Example: **Recall the Bernoulli trial with outcomes H and T. The Bernoulli random variable X takes value 1 at H and value 0 at T. Clearly $\mathbb{E}[X] = p$.**
The Binomial variable Y counts the number of H in n independent Bernoulli trials. Clearly

$$Y = X_1 + X_2 + \cdots + X_n \ ,$$

X_i denoting the Bernoulli random variable of the i^{th} repetition. Therefore the expected number of H in n independent Bernoulli trials with $\mathbb{P}[[\{H\}] = p$ is

$$\sum_{i=1}^{N} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np \ .$$

Theorem: **Let $X : S \to \mathbb{R}$ be a random variable and $\Phi : \mathbb{R} \to \mathbb{R}$ a function. Then**

$$\mathbb{E}[\Phi \circ X] = \sum_{x \in \mathbb{R}} \Phi(x) \cdot \mathbb{P}[X = x] \, .$$

Proof:

$$\begin{aligned} & \mathbb{E}[\Phi \circ X] \\ = & \textstyle\sum_{s \in S} \Phi\big(X(s)\big) \\ = & \textstyle\sum_{x \in \mathbb{R}} \sum_{\{s : X(s) = x\}} \Phi\big(X(s)\big) p(s) \\ = & \textstyle\sum_{x \in \mathbb{R}} \sum_{\{s : X(s) = x\}} \Phi(x) p(s) \\ = & \textstyle\sum_{x \in \mathbb{R}} \Phi(x) \sum_{\{s : X(s) = x\}} p(s) \\ = & \textstyle\sum_{x \in \mathbb{R}} \Phi(x) \mathbb{P}[X = x] \end{aligned}$$

Example: **Let X be a random variable. The sqaure of its deviation from its mean is called its Variance:**

$$\mathbf{var}(X) \stackrel{\text{def}}{=} \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2] - \mu_X^2 \, .$$

The variance of the Bernoulli Variable X above is therefore

$$0^2\mathbb{P}[X = 0] + 1^2\mathbb{P}[X = 1] - p^2 = p(1 - p) \; .$$

Definition: **Two random variables X, Y are independent if for all $x, y \in \mathbb{R}$**

$$\mathbb{P}[X = x \;\wedge\; Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y] \; .$$

Theorem: **If X, Y are independent and $\Phi, \Psi : \mathbb{R} \to \mathbb{R}$ then $\Phi(X), \Psi(Y)$ are independent.**

Theorem: **If X, Y are independent then**

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \; .$$

Proof:

$$
\begin{aligned}
& \mathbb{E}[X \cdot Y] \\
= & \sum_{z \in \mathbb{R}} z \cdot \mathbb{P}[XY = z] \\
= & \sum_{z \in \mathbb{R}_*} z \cdot \sum_{y \in \mathbb{R}_*} \mathbb{P}[XY = z, Y = y] \\
= & \sum_{z \in \mathbb{R}_*} z \cdot \sum_{y \in \mathbb{R}_*} \mathbb{P}[X = z/y, Y = y] \\
= & \sum_{z \in \mathbb{R}_*} z \cdot \sum_{y \in \mathbb{R}_*} \mathbb{P}[X = z/y]\mathbb{P}[Y = y] \\
= & \sum_{y \in \mathbb{R}_*} \mathbb{P}[Y = y] \cdot \sum_{z \in \mathbb{R}_*} z\mathbb{P}[X = z/y]
\end{aligned}
$$

with $z/y = x$**,** $z = xy$**:**

$$
\begin{aligned}
= & \sum_{y \in \mathbb{R}_*} \cdot \sum_{x \in \mathbb{R}_*} xy\mathbb{P}[X = x]\mathbb{P}[Y = y] \\
= & \; y \sum_{y \in \mathbb{R}_*} \mathbb{P}[Y = y] \cdot x \sum_{x \in \mathbb{R}_*} \mathbb{P}[X = x] \\
= & \; \mathbb{E}[Y] \cdot \mathbb{E}[X] \, .
\end{aligned}
$$

Example: **Recall that the Binomial variable Y counts the number of H in n Bernoulli trials in n Bernoulli trials and equals $\sum_{i=1}^{n} X_i$ where X_i denotes the Bernoulli random**

variable of the i^{th} repetition. The latter has expectation p.

$$
\begin{aligned}
\mathbf{var}(Y) &\stackrel{\text{def}}{=} \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i \quad - np\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n}(X_i - p)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i,j=1}^{n}(X_i - p)(X_j - p)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}(X_i - p)(X_i - p)\right] \\
&+ \mathbb{E}\left[\sum_{i \neq j}(X_i - p)(X_j - p)\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[(X_i - p)(X_i - p)] \\
&+ \sum_{i \neq j} \mathbb{E}[(X_i - p)(X_j - p)] \\
&= \sum_{i=1}^{n} \mathbb{E}[(X_i - p)^2] = np(1-p) \ .
\end{aligned}
$$

Theorem: **If $X_1, X_2, \ldots, X_n$ are independent random variables then**

$$\mathbf{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbf{var}(X_i)\ .$$

Theorem – Chebysheff's Inequality: For any random variable X and constant $\lambda > 0$

$$\mathbb{P}[|X| \geq \lambda] \leq \frac{\mathbb{E}[X^2]}{\lambda^2} .$$

Proof: Consider the random variable Y that equals λ on the set $[X \geq \lambda] \subseteq S$ and zero elsewhere: $Y = \lambda I_{[|X| \geq \lambda]}$. Clearly

$$Y^2 \leq X^2 \quad \text{and} \quad \mathbb{E}[Y^2] = \lambda^2 \cdot \mathbb{P}[X \geq \lambda] .$$

Hence $$\lambda^2 \mathbb{P}[X \geq \lambda] = \mathbb{E}[Y^2] \leq \mathbb{E}[X^2]$$

and the claim follows. **QED**

Theorem – The weak law of large numbers: **Let $X_1, X_2, \ldots$ be independent repetitions of the same random variable, and let μ denote the common value of their expectations and v the common value of their variances. Then for all $\epsilon > 0$**

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right] \to 0 \, .$$

Proof: **The random variable**

$$Y \stackrel{\text{def}}{=} \sum_{i=1}^{n} (X_i - \mu)$$

has expectation zero and variance $\text{var}(Y) = \mathbb{E}[Y^2] = nv$. **By Chebysheff's inequality,**

$$\mathbb{P}[|Y/n| \geq \epsilon] = \mathbb{P}[Y \geq n\epsilon] \leq \frac{nv}{n^2\epsilon^2} = \frac{v}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0 \, .$$