

CS 547 Lecture 9: Conditional Probabilities and the Memoryless Property

Daniel Myers

Joint Probabilities

For two events, E and F , the *joint probability*, written $P(EF)$, is the the probability that both events occur.

For example, let E be “the probability that a die roll is even” and F be “the probability that a die roll is greater than 3”. We have the following sets to describe each event:

$$\begin{aligned}E &= \{2, 4, 6\} \\F &= \{4, 5, 6\} \\E \cap F &= \{4, 6\}\end{aligned}$$

The probability that the joint event occurs is the probability that the outcome is in $E \cap F$, which is $\frac{2}{6}$.

Conditional Probabilities

Frequently, we are interested in analyzing the probability of an event with respect to some known information. If E and F are events, the *conditional probability of E given F* is

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Intuitively, $P(EF)$ is the portion of the sample space that is in both E and F . If we know that a point of the sample space is definitely in F , the probability that it is also in E is given by the portion of F that overlaps with E , which is exactly what the formula calculates.

Simply rearranging the formula gives us a way to calculate joint probabilities in terms of conditional probabilities

$$P(EF) = P(E|F)P(F)$$

Independence

Two events are *independent* if $P(E|F) = P(E)$. That is, knowing that F has occurred tells us nothing about whether E also occurs. Using the formula for joint probability, we get

$$P(EF) = P(E|F)P(F) = P(E)P(F)$$

The joint probability of independent events is just the product of their individual probabilities.

Using this knowledge, we’re now able to interpret some earlier results. Recall the pmf of the geometric random variable,

$$f(k) = (1 - p)^{k-1}p$$

We defined this to be the probability that a series of independent Bernoulli trials yields its first success on trial k . That is, we are interested in the probability of a sequence of $k - 1$ failures, each of which occur with probability $1 - p$, and one success, which occurs with probability p . Because the Bernoulli trials are independent, the probability of this sequence is simply the product of the probabilities of its individual events, which gives pmf formula.

Distribution of the Minimum of Exponentials

Suppose we have a set of independent and identically distributed (iid) variables, X_1, X_2, \dots, X_k , which are exponentially distributed with the same parameter λ . What is the distribution of the random variable $X = \min(X_1, X_2, \dots, X_k)$?

Let's reason about $P(X > t)$. X is defined as the minimum of all the X_i , so the only way that X can be greater than a certain value t is if all of its components are also greater than t . Therefore,

$$P(X > t) = P(X_1 > t \text{ and } X_2 > t \text{ and } \dots \text{ and } X_k > t)$$

The individual X_i are all independent, so we can re-write the joint probability as the product of their individual probabilities.

$$P(X > t) = P(X_1 > t) P(X_2 > t) \dots P(X_k > t)$$

To find the final distribution, we need to know $P(X_i > t)$ when $X_i \sim \exp(\lambda)$. This is given by the complementary CDF (CCDF) of the random variable, which is defined in terms of the CDF, $F(t)$, as

$$\begin{aligned}\bar{F}(t) &= P(X_i > t) \\ &= 1 - P(X_i \leq t) \\ &= 1 - F(t)\end{aligned}$$

For the exponential distribution,

$$\begin{aligned}\bar{F}(t) &= 1 - F(t) \\ &= 1 - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t}\end{aligned}$$

Now that we have the exponential CCDF, we can find the distribution of X .

$$\begin{aligned}P(X > t) &= P(X_1 > t) P(X_2 > t) \dots P(X_k > t) \\ &= e^{-\lambda t} e^{-\lambda t} \dots e^{-\lambda t} \\ &= e^{-k\lambda t}\end{aligned}$$

The distribution of the minimum of a set of k iid exponential random variables is also exponentially distributed with parameter $k\lambda$. This result generalizes to the case where the variables are still independent, but have different parameters.

The Memoryless Property

The memoryless property tells us about the conditional behavior of exponential random variables. It's one of our key results, which we'll use in deriving the solution of queueing systems.

Let X be exponentially distributed with parameter λ . Suppose we know $X > t$. What is the probability that X is also greater than some value $s + t$? That is, we want to know

$$P(X > s + t \mid X > t)$$

This type of problem shows up frequently in queueing systems where we're interested in the time between events. For example, suppose that jobs in our system have exponentially distributed service times. If we have a job that's been running for one hour, what's the probability that it will continue to run for more than two hours?

Using the definition of conditional probability, we have

$$P(X > s + t | X > t) = \frac{P(X > s + t \text{ and } X > t)}{P(X > t)}$$

If $X > s + t$, then $X > t$ is redundant, so we can simplify the numerator.

$$P(X > s + t | X > t) = \frac{P(X > s + t)}{P(X > t)}$$

Using the CCDF of the exponential distribution,

$$P(X > s + t | X > t) = \frac{P(X > s + t)}{P(X > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}}$$

The $e^{-\lambda t}$ terms cancel, giving the surprising result

$$P(X > s + t | X > t) = e^{-\lambda s}$$

It turns out that the conditional probability does not depend on t ! The probability of an exponential random variable exceeding the value $s + t$ given t is the same as the variable originally exceeding that value s , regardless of t . In our job example, the probability that a job runs for one additional hour is the same as the probability that it ran for one hour originally, regardless of how long it's been running.

The exponential distribution is *memoryless* because the past has no bearing on its future behavior. Every instant is like the beginning of a new random period, which has the same distribution regardless of how much time has already elapsed.

The exponential is the only memoryless continuous random variable.

Implications of the Memoryless Property

The memoryless property makes it easy to reason about the average behavior of exponentially distributed items in queuing systems.

Suppose we're observing a stream of events with exponentially distributed interarrival times. Because of the memoryless property, the expected time until the next event is always $\frac{1}{\lambda}$, no matter how long we've been waiting for a new arrival to occur.

This behavior is a bit counterintuitive. We might expect that arrivals get *more* likely the longer we wait. For example, if the bus is supposed to come every ten minutes, and we have been waiting for nine minutes without seeing a bus, we expect that the next bus should be along very soon. If the time between bus arrivals is exponentially distributed, however, the memoryless property tells us that our waiting time – no matter how long it's been – is of no use in predicting when the next bus will arrive¹.

Suppose we have a queue with exponentially distributed service times. If a new customer arrives to the queue to find someone in service, the *residual service time* is the time until the currently running customer finishes service and departs the queue. Because of the memoryless property, the distribution of the residual service times does not depend on how long the customer has been in service. The probability that the current customer runs for an additional minute and then departs is the same as the probability that a new customer just entering service runs for one minute. Likewise, the average remaining service time is simply \bar{s} , the expected time for a new customer just entering service.

¹Of course, the time between real bus arrivals is never exponentially distributed. People want to know exactly when their buses will come, so bus schedules are nearly deterministic, which justifies our intuition about longer waiting times increasing the probability of an arrival.

Failure Rates

In the field of reliability theory, it's common to use a random variable to represent the lifespan of a component. One of the main problems in this area is predicting the likelihood that a component fails in the very near future given its current age. This likelihood is summarized by the *failure rate* (also called the hazard rate) of the component².

For most components, the failure rate changes with time. There are three possible relationships.

- increasing failure rate
- decreasing failure rate
- constant failure rate

If the failure rate is *increasing*, then failures become more likely as the component ages. Most manufactured products behave in this way – they're built to last for a certain amount of time, then fail. Failures become more likely as the product wears out and the end of its lifespan approaches.

A *decreasing* failure rate implies that the probability of failure decreases with the passage of time. In other words, the longer a component has worked, the more likely it is to continue working. UNIX processes have been shown to have decreasing failure rates – the longer a job runs, the more likely it is to continue running³. Human lifespans also have decreasing failure rate, as surviving further into adulthood makes it more likely that you will live to old age, at least until you reach the upper limit of your natural lifespan.

The final category corresponds to components with exponentially distributed lifespans. Because of the memoryless property, the length of time a component has functioned in the past has no bearing on its future behavior, so the probability that the component fails in the near future is always the same and doesn't depend on its current age.

²It's possible to formally derive a function for the failure rate in terms of the lifespan distribution. The actual function isn't directly relevant to our work, so I've chosen not to present it here.

³This justifies the use of schedulers that steadily decrease the priority of a job as it runs. More on this in future lectures.