

Session 5: Named Models and Extremes

Movement II: One Random Variable

Instructor: John Lipor

1 Launch

Session 4 built random variables from scratch, one ad-hoc count at a time. But the same few shapes keep reappearing, and three questions from a single search make the point. *How many of 25 cells produce a detection on one sweep?* Each cell succeeds or fails on its own, a fixed number of independent yes/no trials: that count is *binomial*. *How many passes until this cell finally gives up its target?* We repeat one trial until the first success: that waiting time is *geometric*. *If two vehicles each search their own cell, how long until the **whole** job is done?* That is the *maximum* of two waiting times, and here our tools run out: we have no formula for a maximum. Today we catalog the handful of named models that cover almost every sensor and count in the course, and we forge the one genuinely new tool of the session, the CDF method, which answers every question about a maximum or a minimum.

2 Development

2.1 Bernoulli and binomial

The atom of every count is a single yes/no trial.

Definition 1 (Bernoulli). $X \sim \text{Bernoulli}(p)$ takes the value 1 with probability p and 0 with probability $1 - p$. It is the indicator of a success, so $\mathbb{E}[X] = p$ and $\text{Var}(X) = p(1 - p)$.

Stacking n independent copies of the same trial and counting the successes gives the workhorse count model.

Definition 2 (Binomial). $X \sim \text{Binomial}(n, p)$ is the number of successes in n independent trials, each succeeding with probability p . Its PMF is

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n, \quad (1)$$

the factor $\binom{n}{k}$ counting which trials succeeded.

We do not need the PMF to find the mean and variance: Session 4’s machinery delivers them for free. Write $X = \sum_{i=1}^n X_i$ with $X_i \sim \text{Bernoulli}(p)$ the indicator of trial i . By linearity (no independence needed),

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = np. \quad (2)$$

Because the trials are independent, the cross terms in the variance of the sum vanish and the variances add:

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p). \quad (3)$$

This is the whole point of the indicator method: the binomial is not a new object, it is n Bernoullis added up, and its summary statistics follow from linearity and independence exactly as in the last session.

2.2 Geometric: waiting for the first success

Instead of fixing the number of trials, fix the goal (the first success) and let the number of trials be random.

Definition 3 (Geometric). $X \sim \text{Geometric}(q)$ is the number of independent trials up to and including the first success, where each trial succeeds with probability q . Its support is $\{1, 2, 3, \dots\}$ and

$$p_X(k) = q(1-q)^{k-1}, \quad k = 1, 2, 3, \dots, \quad (4)$$

since the first $k-1$ trials must fail and the k -th must succeed.

Convention (fix this once and never deviate). We count *the trial on which the first success occurs*, so the geometric starts at 1. Many textbooks instead count the number of *failures before* the first success, which starts at 0; the two conventions differ by exactly 1. Ours is the “number of passes until this cell is searched successfully” count, which is what a search log records.

The survival probability follows directly: the event $\{X > k\}$ is exactly “the first k trials all failed,” so

$$\mathbb{P}(X > k) = (1-q)^k, \quad k = 0, 1, 2, \dots \quad (5)$$

For the mean we introduce a general tool that will pay off repeatedly.

Theorem 1 (Tail-sum formula). *For a nonnegative integer-valued random variable X ,*

$$\mathbb{E}[X] = \sum_{k \geq 1} \mathbb{P}(X \geq k). \quad (6)$$

Proof: HW2, Problem 1(b). We use only the statement here. Applied to the geometric, $\mathbb{P}(X \geq k) = \mathbb{P}(X > k-1) = (1-q)^{k-1}$ by (5), so the mean is a geometric series:

$$\mathbb{E}[X] = \sum_{k \geq 1} (1-q)^{k-1} = \frac{1}{1-(1-q)} = \frac{1}{q}. \quad (7)$$

A cell searched with per-pass success probability q takes, on average, $1/q$ passes to give up its target. The rarer the success, the longer the wait, in exact inverse proportion.

2.3 Memorylessness

The geometric has a startling property: having waited without success tells you nothing about how much longer you must wait.

Proposition 1 (Memorylessness). *If $X \sim \text{Geometric}(q)$, then for all nonnegative integers m, k ,*

$$\mathbb{P}(X > m + k \mid X > m) = \mathbb{P}(X > k). \quad (8)$$

Proof. Using the definition of conditional probability and the survival formula (5), and noting $\{X > m + k\} \subseteq \{X > m\}$ so the intersection is $\{X > m + k\}$,

$$\mathbb{P}(X > m + k \mid X > m) = \frac{\mathbb{P}(X > m + k)}{\mathbb{P}(X > m)} = \frac{(1 - q)^{m+k}}{(1 - q)^m} = (1 - q)^k = \mathbb{P}(X > k). \quad (9)$$

□

In search terms: a cell that has resisted m passes is, *given the target is still there*, exactly as hard to clear as it was on the first pass. The trials do not tire and do not learn; each pass faces the same odds.

This is easy to misread, and the misreading collides with everything Movement I taught. Memorylessness is a statement about the *conditional* law of the waiting time *given that the target is present in the cell*. It says nothing about whether the target *is* present. Those are different questions, and Session 2 already answered the second one: each failed pass is evidence, and the failed-search update flows probability out of the searched cell, so unconditionally our belief that the target is even in this cell *falls* with every empty pass. Both facts hold at once, and they do not conflict. Conditioned on presence, the remaining wait is undiminished; but our confidence in that very conditioning erodes. Persistence that ignores the second fact, and searches forever because “the odds each pass are unchanged,” is exactly the error Session 2 warned against.

Two clocks, not one. Memorylessness governs the wait *given presence*; Bayes governs *belief in presence*. A run of failures leaves the first clock untouched and runs the second one down. Never let “the geometric is memoryless” become “the failures told us nothing”: they told us nothing about the wait, and a great deal about the location.

2.4 Poisson: the model for rare counts

Definition 4 (Poisson). $X \sim \text{Poisson}(\lambda)$ has support $\{0, 1, 2, \dots\}$ and PMF

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots \quad (10)$$

Its mean and variance are both λ (the mean is a short exercise from the series for e^λ ; the variance is left as one).

The Poisson earns its central place as the limit of a binomial with many trials, each rarely successful. Fix a count k and let $n \rightarrow \infty$ while holding the mean $np = \lambda$ fixed, i.e., $p = \lambda/n$:

$$\begin{aligned} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} &= \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\rightarrow 1} \cdot \frac{\lambda^k}{k!} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} \\ &\longrightarrow \frac{\lambda^k e^{-\lambda}}{k!}. \end{aligned} \quad (11)$$

The first factor is a product of k ratios each tending to 1, and the third is the classical limit $(1 - \lambda/n)^n \rightarrow e^{-\lambda}$. This is the *law of rare events*: a great many independent chances, each unlikely, produce a Poisson count.

The limit is not a curiosity. It says the Poisson is what a binomial *becomes* whenever the batch is large and each success is rare, which is exactly the regime of most sensing counts.

Poisson is the universal rare-count model. Photons striking a detector in a gate, packets arriving at a relay, false alarms across a large sensor array: each is a huge number of independent, individually unlikely opportunities. That is why Poisson is the count model of choice, and why it returns in Session 10 as the noise model for photon-counting detection.

2.5 The CDF method for extremes

Now to the tool the Launch promised. When several independent quantities are in play and we ask for the largest or the smallest, neither the PMF nor the PDF is the natural object; the CDF is. The reason is that “ $\max \leq x$ ” and “ $\min > x$ ” both factor cleanly across independent variables, whereas “ $\max = x$ ” does not.

Theorem 2 (Distributions of a maximum and a minimum). *Let $X_1, \dots, X_m$ be independent and identically distributed with common CDF F . Write $M = \max_i X_i$ and $L = \min_i X_i$. Then*

$$\mathbb{P}(M \leq x) = F(x)^m, \quad \mathbb{P}(L > x) = (1 - F(x))^m. \quad (12)$$

Proof. The maximum is at most x exactly when *every* X_i is at most x :

$$\mathbb{P}(M \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_m \leq x) = \prod_{i=1}^m \mathbb{P}(X_i \leq x) = F(x)^m, \quad (13)$$

using independence to factor and identical distribution to collapse each factor to $F(x)$. Symmetrically, the minimum exceeds x exactly when *every* X_i exceeds x :

$$\mathbb{P}(L > x) = \mathbb{P}(X_1 > x, \dots, X_m > x) = \prod_{i=1}^m \mathbb{P}(X_i > x) = (1 - F(x))^m. \quad (14)$$

In each case independence turned the joint statement about all m variables into a product, and identical distribution collapsed the factors. $\square$

The two formulas share one moral, worth stating on its own before we use them.

When you hear “max” or “min,” reach for the CDF. The event $\{\max \leq x\}$ and the event $\{\min > x\}$ each say something about *all* the variables at once, and independence turns “all at once” into a product. Never try to attack a maximum or a minimum through its PMF or density directly; write down its CDF (or its survival probability) and differentiate only at the end if you need a density.

3 Worked examples

3.1 (a) When is the job finished? The maximum of two geometrics

The Launch asked when a two-cell job is finished. Each of two cells is known to hold a target, and a dedicated vehicle searches each one, the passes synchronized; on every pass each vehicle independently detects its target with probability $q = 0.5$. Vehicle i ’s time to detection is $X_i \sim \text{Geometric}(0.5)$, and the job is done at $M = \max(X_1, X_2)$.

Distribution. Reach for the CDF. Each X_i has $F(k) = \mathbb{P}(X_i \leq k) = 1 - 0.5^k$ by (5), so the maximum half of Theorem 2 gives

$$\mathbb{P}(M \leq k) = (1 - 0.5^k)^2, \quad \text{e.g.} \quad \mathbb{P}(M \leq 3) = \left(\frac{7}{8}\right)^2 = \frac{49}{64} \approx 0.766. \quad (15)$$

Three passes finish the job only about three times in four, even though each single target is found within three passes with probability $7/8$.

Mean. The tail-sum formula (Theorem 1) asks for $\mathbb{P}(M \geq k) = \mathbb{P}(M > k - 1)$, so work with the survival probability

$$\mathbb{P}(M > j) = 1 - (1 - 0.5^j)^2 = 2(0.5)^j - (0.25)^j, \quad j = 0, 1, 2, \dots \quad (16)$$

Reindexing with $j = k - 1$ and summing two geometric series,

$$\mathbb{E}[M] = \sum_{j \geq 0} \mathbb{P}(M > j) = 2 \sum_{j \geq 0} (0.5)^j - \sum_{j \geq 0} (0.25)^j = 2 \cdot 2 - \frac{4}{3} = \frac{8}{3} \approx 2.67 \text{ passes.} \quad (17)$$

A single cell takes $1/q = 2$ passes on average. Two cells searched in parallel take longer, because we must wait for the slower search, but far less than twice as long.

The job has memory. Each search is memoryless, but the job is not. The chance of finishing on the next pass, given that the job is not yet done, is $\mathbb{P}(M = k + 1 \mid M > k) = 1 - \mathbb{P}(M > k + 1) / \mathbb{P}(M > k)$, which equals $1/4$ on the first pass (both targets must be found at once), $5/12 \approx 0.417$ on the second, and climbs toward $1/2$ thereafter: the longer the job runs, the more likely one target is already found, leaving a single geometric search. Proposition 1 is a property of one geometric clock, not of a maximum of them.

3.2 (b) A binomial that is really a Poisson, and one that is not

A large array has $n = 1000$ cells, and each independently raises a *false* alarm (a detection where nothing is present) with probability 0.003 on a sweep. Let C be the number of false

alarms. Exactly, $C \sim \text{Binomial}(1000, 0.003)$, with mean $np = 3$. Because n is large and p is small with np moderate, the law of rare events says C is very nearly $\text{Poisson}(3)$.

Compare the two at $k = 0$, the probability of a clean sweep:

$$\mathbb{P}(C = 0)_{\text{exact}} = (1 - 0.003)^{1000} = 0.997^{1000} \approx 0.0496, \quad (18)$$

$$\mathbb{P}(C = 0)_{\text{Poisson}} = e^{-3} \approx 0.0498. \quad (19)$$

The two agree to three decimal places. Is it the large n that makes this work? Try a binomial with the same mean but a few likely trials: $n = 10$ cells with $p = 0.3$, so again $np = 3$. Now $\mathbb{P}(C = 0) = 0.7^{10} \approx 0.0282$, while the Poisson approximation still says $e^{-3} \approx 0.0498$, almost twice too large. Matching the mean does not make a count Poisson.

The limit computation tells us what does. At $k = 0$ the exact value is $(1 - p)^n = \exp(n \ln(1 - p))$, and the series $\ln(1 - p) = -p - p^2/2 - p^3/3 - \dots$ (valid for $0 \leq p < 1$) gives, with $\lambda = np$,

$$(1 - p)^n = e^{-\lambda} \cdot \exp\left(-\lambda\left(\frac{p}{2} + \frac{p^2}{3} + \dots\right)\right). \quad (20)$$

The first factor is the Poisson value; the second is the error, and with λ fixed it is controlled by p , not by n . For $p = 0.003$ it is about $e^{-\lambda p/2} = e^{-0.0045} \approx 0.9955$, which is exactly the ratio of the two unrounded values above. For $p = 0.3$ the correction terms are no longer small, and the exact value is only about 0.57 of the Poisson one. The law of rare events is a statement about *rare*: the Poisson model is accurate when each individual chance is unlikely, and then it needs only the single number λ , no factorials and no thousand-fold product.

4 Consolidation

Work these in class.

1. (*Mechanical, geometric.*) A cell is searched with per-pass success probability $q = 0.25$. What is the expected number of passes to detection, and what is the probability the cell survives more than 4 passes?
Answer: $\mathbb{E}[X] = 1/0.25 = 4$ passes; $\mathbb{P}(X > 4) = (0.75)^4 = 0.3164$.
2. (*CDF method.*) Two independent fair dice are rolled, each with CDF $F(k) = k/6$ for $k = 1, \dots, 6$. Using the CDF method, find $\mathbb{P}(\max \leq 4)$ and $\mathbb{P}(\min \geq 5)$.
Answer: $\mathbb{P}(\max \leq 4) = F(4)^2 = (4/6)^2 = 4/9$; and $\mathbb{P}(\min \geq 5) = \mathbb{P}(\min > 4) = (1 - F(4))^2 = (2/6)^2 = 1/9$.
3. (*Conceptual, memorylessness.*) A cell has been searched 10 times without success. A colleague argues: “the geometric is memoryless, so those failures carry no information; our belief the target is here is unchanged.” Identify the one true clause and the one false clause.
Answer: True: given the target is present, the remaining wait is unchanged (memorylessness of the conditional law). False: our belief the target is present is unchanged; Session 2’s failed-search update has driven that belief down with every empty pass.

5 Synthesis

The course thesis asks first, “what is uncertain?” This session’s answer is that most of the time the uncertain count has one of a few recognizable shapes: a fixed batch of independent trials is binomial, a wait for the first success is geometric, and a flood of individually unlikely chances is Poisson. Naming the shape hands you its mean, its spread, and its behavior at a glance. Two ideas reached past the catalog. Memorylessness split one clock into two, the undiminished wait *given* presence and the eroding belief *in* presence, and reconciled them with Movement I rather than contradicting it. And the CDF method answered the questions the PMF could not: whenever “max” or “min” is spoken, the CDF factors across independent variables and the PMF does not, so the CDF is the tool. Next session carries the CDF into continuous territory, where variables have no PMF at all but every one of them still has a CDF, and where the geometric’s continuous twin, the memoryless waiting time on a clock rather than a counter, appears as the exponential distribution.

Self-check questions

1. You need the mean of a $\text{Geometric}(q)$ waiting time and you know only its survival probability $\mathbb{P}(X > k) = (1 - q)^k$. Which formula turns the survival probabilities into the mean, and what does it give?

Answer: the tail-sum formula $\mathbb{E}[X] = \sum_{k \geq 1} \mathbb{P}(X \geq k) = \sum_{k \geq 1} (1 - q)^{k-1} = 1/q$.

2. Five independent sensors have iid detection times with common CDF F . In one line, what is the probability the *first* detection happens after time x , and why is this easier than working with the PMF of the minimum?

Answer: $\mathbb{P}(\min > x) = (1 - F(x))^5$; the event “all five exceed x ” factors across independent sensors, whereas “the minimum equals x ” does not.

3. A cell has failed 8 searches, each with per-pass success probability q . A student says the geometric’s memorylessness means the 8 failures are uninformative. What, precisely, is unchanged, and what has in fact changed?

Answer: unchanged is the remaining wait *given the target is present*; changed (fallen) is the belief that the target is present at all, by the Session 2 failed-search update.