Tips & Tricks in Probability

Sinho Chewi, Alvin Wan

Fall 2016

University of California, Berkeley

Symmetry

The principle behind symmetry is that if two random variables are exchangeable, that is, they can be swapped without changing the scenario, then they have the same distribution.

You have $n - 1$ friends and a box containing n chocolates. One of the chocolates is dark chocolate, and you want it for yourself. Each person picks a chocolate without looking. Do you want to be the first or the last person to pick a chocolate?

It doesn't matter! Think about the order in which chocolates are drawn as a random permutation. By definition, a random permutation assigns equal probability for the dark chocolate to each position.

Symmetry

Pick n points i.i.d. $U[0,1]$. The n points split the interval $[0,1]$ into $n+1$ different pieces. By symmetry, the expected length of each segment is the same. Consequently, the minimum of the points has expectation $1/(n+1)$, the next smallest has expectation $2/(n+1)$, and so on, until the maximum has expectation $n/(n+1)$.

Pick 3 points randomly on the circumference of a 24-hour clock. These points represent the arrival times of three different buses throughout the day. You are at a bus stop, and you have just watched a bus leave without you. What is the expected time until the next bus arrives? 8 hours.

More Symmetry

Let X_1, X_2, X_3 be i.i.d. and write $S = X_1 + X_2 + X_3$. What is $E[X_1 + X_2 \mid S]$?

By symmetry: $2S/3$. By a more formal argument: (X_1, X_2) has the same distribution as (X_1, X_3) and (X_2, X_3). Therefore,

$$E[(X_1 + X_2) + (X_1 + X_3) + (X_2 + X_3) \mid S] = 2S$$

so each of the three terms is $2S/3$. We did not need to know the distribution of X_i.

Minimum, Maximum, Tail Sum

Suppose that $X_1, \ldots, X_n$ are i.i.d.

$$P(\min\{X_1, \ldots, X_n\} > x) = P(X_1 > x, \ldots, X_n > x) = P(X_i > x)^n$$
$$P(\max\{X_1, \ldots, X_n\} < x) = P(X_1 < x, \ldots, X_n < x) = P(X_i < x)^n$$

Example: Roll an n-sided die twice. What is the expectation of the maximum of the rolls?

Let X_i be the outcome of roll i and $X = \max\{X_1, X_2\}$.

$$P(X \geq x) = 1 - P(X < x) = 1 - P(X_i < x)^2 = 1 - \left(\frac{x-1}{n}\right)^2$$

Now, use the tail sum formula.

$$E[X] = \sum_{x=1}^{n} P(X \geq x) = \sum_{x=1}^{n} \left(1 - \left(\frac{x-1}{n}\right)^2\right)$$

Indicators

Indicators are used for random variables that count quantities. "Compute $E[X]$, where X is the number of floors..." Define an indicator for each floor!

Remember: $E[1_A] = P(A)$. The expectation of an indicator is the probability of the event that it indicates! To compute $\mathrm{var}(X)$, calculate $E[X^2]$ using:

$$E[X^2] = \underbrace{\sum_{i=1}^{n} E[X_i^2]}_{n \text{ terms}} + \underbrace{\sum_{i \neq j} E[X_i X_j]}_{n(n-1) \text{ terms}} = nP(A_i) + n(n-1)P(A_i \cap A_j)$$

...unless your indicators are *independent*, in which case $\mathrm{var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{var}(X_i)$.

Indicators: A Clever Application

We have r red marbles and b blue marbles in a bag. Sampling *without* replacement, what is the expected number of blue marbles we draw before the kth red marble? (*Remark*: If we sample *with* replacement, use conditional expectation.)

Define an indicator for each blue marble: X_i indicates that the ith blue marble is drawn *before the kth red marble*. What is $E[X_i]$? $k/(r+1)$ (think about laying the r red marbles and the ith blue marble in a line). Therefore, our answer is $bk/(r+1)$.

Indicators: Another Application

You are hosting a party and you invite your friends. For $0 \le i \le k$, n_i is the number of your friends who have exactly i children. Each friend attends independently, with probability p, and if your friend attends, he/she brings all of his/her children. Let X be the total number of attendees. $E[X]$? $\mathrm{var}(X)$?

Actually, this is a scaled version of the binomial distribution: for group i, the number of attendees is binomial with n_i trials, probability of success p, and *scaled* by $i+1$.

$$E[X] = \sum_{i=0}^{k} (i+1) E[X_i] = \sum_{i=0}^{k} (i+1) n_i p$$

$$\mathrm{var}(X) = \sum_{i=0}^{k} (i+1)^2 \,\mathrm{var}(X_i) = \sum_{i=0}^{k} (i+1)^2 n_i p (1-p)$$

Moral of the story: always look for the easier way.

Complements of Events

Often, it is much easier to calculate the probability of the *complement* of an event. Standard example: what is the probability that at least one condition is met? Instead, consider the complement: what is the probability that none of the conditions are met?

Throw n balls in m boxes. What is the variance of the number of non-empty boxes?

Define an indicator for each box. What is the probability that a box is non-empty? What is the probability that boxes i and j are non-empty? Easier question: what is the probability that boxes i and j are both empty?

$$\mathrm{var}(\text{non-empty}) = \mathrm{var}(m - \text{non-empty}) = \mathrm{var}(\text{empty})$$

Iterated Expectation

- We use the **Law of Iterated Expectation** when we *wish we had more information.*

$$E[X] = E[E[X \mid Y]]$$

- It's easy to compute $E[X]$, when $X \sim \textsc{Bin}(n, p)$ for constants n, p.

$$E[X] = np$$

- What if p is not a constant? Say, $X \sim \textsc{Bin}(n, P)$, $P \sim U[0, 1]$, for constant n? We *wish we knew* P, so condition on it!

$$E[X] = E[E[X \mid P]] = E[nP] = nE[P] = \frac{n}{2}$$

Iterated Expectation: Example 1

Q: Consider a six-sided die. How many evens do we expect to roll before rolling our first odd?

Iterated Expectation: Solution 1

Q: Consider a six-sided die. How many evens do we expect to roll before rolling our first odd?

A:

- X = number of evens, Y = rolls needed to see first odd.
- If we have not yet rolled an odd, all of our rolls are even!
  - $X \sim \textsc{Bin}(Y-1, 1)$
  - $Y \sim \textsc{Geom}(1/2)$

$$E[X] = E[E[X \mid Y]] = E[Y-1] = 1$$

Our answer is 1.

Q: Consider a six-sided die. How many twos do we expect to roll before rolling our third three?

Iterated Expectation: Solution 1

Q: Consider a six-sided die. How many twos do we expect to roll before rolling our third three?

A: First, consider just the number of twos we expect before the first three.

- $X =$ number of twos, $Y =$ rolls before the first three.
- If we have not yet rolled a three, we have a 1/5 chance of rolling a two.
  - $X \sim \textsc{Bin}(Y-1, 1/5)$
  - $Y \sim \textsc{Geom}(1/6)$

$$E[X] = E[E[X \mid Y]] = E[(Y-1)/5] = 1$$

This is the same for each three. Our answer is 3.

When solving problems, sometimes the expectation of the $(t+1)$th state can be expressed in terms of the expectation of the tth state, for constants α, β:

- $X(t+1) = \alpha X(t)$
- Solution: $X(t) = \alpha^t X(0)$
- $X(t+1) = \alpha X(t) + \beta$
- Solution: $X(t) = \alpha^t X(0) + \beta \dfrac{1-\alpha^t}{1-\alpha}$

Iterated Expectation: Problem 2a

Q: There are 600 ungraded exams. Each minute, Allen picks up an exam. With probability p, he grades it. He always puts the exam back. How many ungraded exams are there at time $i+1$, T_{i+1}, as a function of the ungraded exams at time i, T_i?

Iterated Expectation: Solution 2a

Q: There are 600 ungraded exams. Each minute, Allen picks up an exam. With probability p, he grades it. He always puts the exam back. How many ungraded exams are there at time $i+1$, T_{i+1}, as a function of the ungraded exams at time i, T_i?

A: With probability $1 - T_i/600$, Allen picks up a graded exam, in which case $T_{i+1} = T_i$. Otherwise, he picks up an ungraded exam and with probability p, he will grade it.

$$
\begin{aligned}
E[T_{i+1} \mid T_i] &= \left(1 - \frac{T_i}{600}\right) T_i + \frac{T_i}{600}(T_i - p) = T_i - \frac{p}{600} T_i \\
&= \left(1 - \frac{p}{600}\right) T_i \\
E[T_{i+1}] &= \left(1 - \frac{p}{600}\right) E[T_i]
\end{aligned}
$$

Iterated Expectation: Problem 2b

Q: There are 600 ungraded exams. Each minute, Allen picks up an exam. With probability p, he grades it. He always puts the exam back. How many exams do we expect Allen to have graded by time t?

Iterated Expectation: Solution 2b

Q: There are 600 ungraded exams. Each minute, Allen picks up an exam. With probability p, he grades it. He always puts the exam back. How many exams do we expect Allen to have graded by time t?

A: We can recognize that this is in the form $X(t+1) = \alpha X(t)$, where $\alpha = 1 - p/600$. We plug our value of p in to our closed-form solution. Keeping in mind that $E[T_0] = 600$,

$$E[T_t] = \alpha^t E[T_0] = 600\left(1 - \frac{p}{600}\right)^t$$

Memoryless Property

The past does not affect the future. Take the current time a, and consider some *additional* time b. The memoryless property states:

$$P(t > a + b \mid t > a) = P(t > b)$$

Light bulbs have i.i.d. lifetimes, $\text{Expo}(\lambda)$. Replace a light bulb as soon as the old one dies. Your friend walks into the room at a random time. What is the expected lifetime of the light bulb that is burning when he walks in? (*Hint*: The answer is not $1/\lambda$.)

By the memoryless property, we expect the light bulb to burn for an additional time $1/\lambda$. By symmetric logic, we expect the light bulb to have already been burning for $1/\lambda$. The expected lifetime is therefore $2/\lambda$. (This is sometimes known as the Inspector's Paradox.)

Summary

- Take advantage of symmetry to simplify calculations.
    - box of chocolates, segments, bus arrivals
- When working with the minimum/maximum of random variables, consider the CDF instead.
    - maximum of two rolls
- Indicators count successes. Use them liberally.
    - marbles without replacement, party with friends

Summary

- Consider the complement of events, because their probabilities might be easier to compute.
  - non-empty boxes
- Condition when you wish you had more information.
  - r.v.s as parameters, number of twos before 3 threes
- You can ignore an exponentially-distributed random variable's past.
  - light bulbs