Probability Theory

Textbook: *Introduction to Probability* by Blitzstein and Hwang

---

Previous Lecture

- CDFs and Valid CDFs
- Functions of rvs
- Indep of rvs

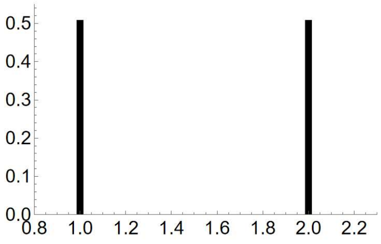
---

§4.1 - Definition of Expectation

Given some random process, what's the "long-run average" of the possible outcomes.

Ex (**Coin Expectation**): Imagine flipping a coin, where $X(H) = 1$ for heads and $X(T) = 2$ for tails.

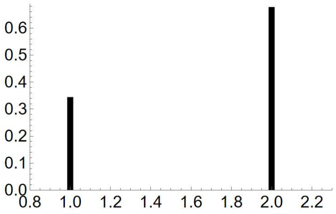
If the probability of each outcome is equal, we would have:



PF. Average is 1.5

Average Value $= \frac{1+2}{2} = \frac{3}{2}$. But what if tails was twice as likely as heads?

Notice that the last calculation can be written $\frac{1+2}{2} = \frac{1}{2} + \frac{2}{2} = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right)$. (each value multiplied by likelihood of that value)



PF. Average is $\frac{5}{3} = 1.6667$

So, *Average* $= 1\left(\frac{1}{3}\right) + 2\left(\frac{2}{3}\right) = \frac{5}{3}$.

Def (**Expectation of a Discrete Rv**): The expected value (also called the expectation or mean) of a discrete X whose distinct possible values are $x_1, x_2, \ldots$ is: $E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j)$. (AKA weighted average).

If support is finite, then we write: $E(X) = \Sigma_x \underbrace{x}_{\text{value}} \underbrace{P(X = x)}_{\text{PMF at } x}$, where the sum ($\Sigma$) is over the support of X.

Ex: Let X be result of rolling a fair 6-sided die, so $X \in \{1,2,3,4,5,6\}$ w/equal probabilities. What's $E(X)$?

$$E(X) = \Sigma_x x P(X = x)$$

$$= 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right) = \tfrac{7}{2} = 3.5$$

Ex: Let $X \sim Bern(p)$ and $q = 1 - p$. What's $E(X)$?

$$E(X) = \Sigma_x x P(X = x) = 0 \cdot q + 1 \cdot p = p.$$

Ex: Define X w/PF: $f_X(x) = \begin{cases} \frac{|x|}{32} & \text{for } x = -10, -5, -1, 0, 1, 5, 10 \\ 0 & \text{otherwise.} \end{cases}$ What's $E(X)$?

$$E(X) = \Sigma_x x P(X = x)$$

$$= -10\left(\tfrac{10}{32}\right) - 5\left(\tfrac{5}{32}\right) - 1\left(\tfrac{1}{32}\right) + 0\left(\tfrac{0}{32}\right) + 1\left(\tfrac{1}{32}\right) + 5\left(\tfrac{5}{32}\right) + 10\left(\tfrac{10}{32}\right) = 0.$$

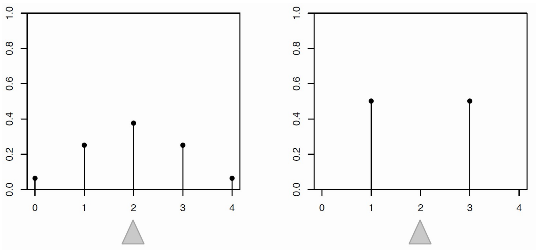
Expectation is undefined if $\sum_{j=1}^{\infty} |x_j| P(X = x_j)$ diverges.

Ex (∞ payout): A game where you pay \$4 to toss a fair coin until you get heads. If heads appears on the n-th toss, you win 2^n dollars. Expected payout: \$$\infty$.

Proposition. If X and Y are discrete w/the same distr, then $E(X) = E(Y)$ (if either side exists).

Proof. In the definition of $E(X)$, we only need to know the PMF of X. ■

☣ X is a function, while $E(X)$ is a constant!



Different PFs can have same $E(X)$

Harvard Video: youtube.com/watch?v=LX2q356N2rU&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=10

---

§4.2 - Linearity of Expectation

Thm (**Linearity of Expectation**): For any X and Y and any constant c,
$E(X+Y) = E(X) + E(Y)$, and $E(cX) = cE(X)$.

Ex: If $E(X) = 5$ and $E(X^2) = 12$, then what is $E(2-7X)$? How about $E(3X^2+4X-2)$?

Ex (**Binomial Expectation**): For $X \sim Bin(n,p)$, let's find $E(X)$ in two ways. First, by definition of expectation,

we have: $E(X) = \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k}$. (Our goal? Get rid of the Σ, turn into $Bin(?,?)$)

Recall that in general: $k\binom{n}{k} = n\binom{n-1}{k-1}$. (team captain)

so $\sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k} = n\sum_{k=0}^{n}\binom{n-1}{k-1}p^k q^{n-k}$

$= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{n-k}$ (pull out p to better match Bin form, and $\binom{n-1}{-1} = 0$ so $k=1$)

$= np\sum_{j=0}^{m}\binom{m}{j}p^j q^{m-j}$ (let $(k-1) \to j$ and $n-1 \to m$)

$= np.$ (since the sum is summing the probabilities of $Bin(m,p)$, which must equal 1)

Therefore, $E(X) = np$.
Using linearity of expectation gives a much shorter proof.

Instead: let's write X as the sum of n indep $Bern(p)$ rvs: $X = I_1 + \ldots + I_n$, where each I_j has expectation:

$E(I_j) = 1p + 0q = p.$

By linearity, $E(X) = E(I_1) + \ldots + E(I_n) = np.$ $\square$

Harvard Video: youtube.com/watch?v=P1fSFvhPf7Q&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=11

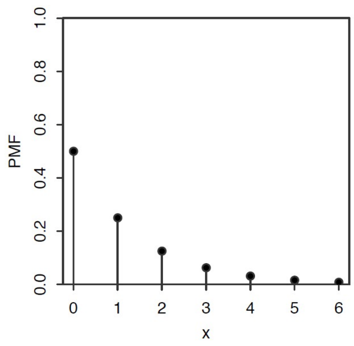
---

§4.3 - Geometric and Negative Binomial

Story (**Geometric Distr**): Consider a sequence of indep Bernoulli trials, each w/ $p \in (0,1)$,
w/trials performed until a success occurs. Let X be the # of failures before the first successful trial.
Then X has Geometric distr w/parameter p. We denote this $X \sim Geom(p)$.

Ex: if we flip a fair coin until it lands Heads for the first time, then "# of Tails before the first occurrence of Heads" is distributed as $Geom(\frac{1}{2})$.

Thm (**Geometric PF**): If $X \sim Geom(p)$, then the PF of X is $P(X = k) = q^k p$ for $k = 0, 1, 2, \ldots$ where $q = 1 - p$.



$Geom(\frac{1}{2})$

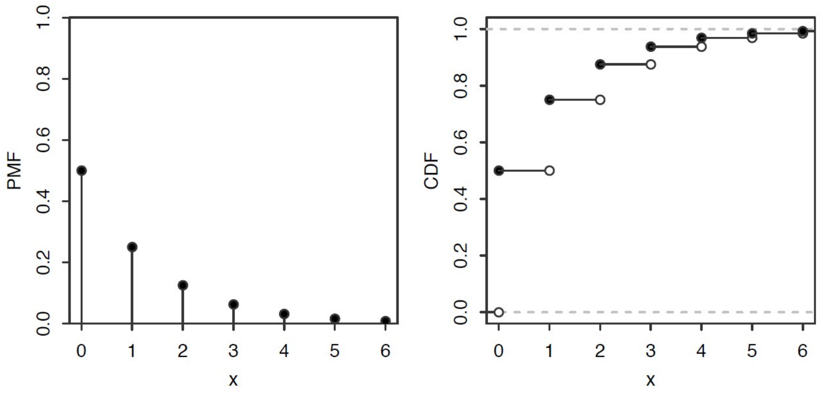
Ex (**Bike Riders**): Assume 5% of people traveling down a sidewalk are on bikes. You watch as folks go by. Calculate the prob that the 8th person will be the 1st bike rider.

Solution: $p = 0.05$. We want 7 non-cyclists, followed by a bike rider:

$$P(X = 7) = (1 - p)^7 p = (0.95)^7(0.05) \approx 0.035 \text{ or } 3.5\%. \qquad \square$$

Thm (**Geometric CDF**): If $X \sim Geom(p)$, then the CDF of X is:

$$F(x) = \begin{cases} 1 - q^{\lfloor x \rfloor + 1}, \text{ if } x \geq 0; \\ 0, \text{ if } x < 0, \end{cases}$$ where $q = 1 - p$ and $\lfloor x \rfloor$ is greatest integer which is less than or equal to x.



$Geom(\frac{1}{2})$

Def (**First Success Distr**). In a sequence of indep Bernoulli *trials* with success prob p, let Y be # of trials until the first successful trial, **including the success**. Then Y has the *First Success distr* w/parameter p; we denote this by $Y \sim FS(p)$.

Converting between $Geom(p)$ **and** $FS(p)$:

If $Y \sim FS(p)$, then $Y - 1 \sim Geom(p)$. We convert between PFs of Y and $Y - 1$ with: $P(Y = k) = P(Y - 1 = k - 1)$.

If $X \sim Geom(p)$, then $X+1 \sim FS(p)$.

Ex (**Geometric Expectation**): Let $X \sim Geom(p)$. Show that $E(X) = \frac{q}{p} = \frac{1-p}{p}$. (useful in HW)

By definition, $E(X) = \sum_{k=0}^{\infty} kq^k p$.

Not a geometric series because of k multiplying each term.

But notice that each term looks similar to kq^{k-1}, the derivative of q^k (w/respect to q), so let's start there. We know:
$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$.

This geometric series converges since $0 < q < 1$. Differentiating both sides: $\sum_{k=0}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$.

If we multiply both sides by pq, we recover the original sum we wanted to find:
$E(X) = \sum_{k=0}^{\infty} kq^k p = \frac{q}{p}$. □

Now let's generalize the geometric distr. Instead of waiting for 1 success, we can wait for any # (r) of successes.

Story (**Negative Binomial Distr**): In a sequence of indep Bernoulli trials w/success prob p, if X is the # of failures before the rth success, then X has the *Negative Binomial distr* w/parameters r and p, denoted $X \sim NBin(r,p)$.

Both Binomial and Negative Binomial distr's are based on indep Bernoulli trials; they differ in the stopping rule and in what they're counting.

- ♦ Binomial counts # of *successes* in a fixed # of *trials*;
- ♦ Negative Binomial counts # of *failures* until a fixed number of *successes*.

Thm (**Negative Binomial PF**): If $X \sim NBin(r,p)$, then the PF of X is $P(X = n) = \binom{n+r-1}{r-1} p^r q^n$ for $n = 0, 1, 2, \ldots$ where $q = 1 - p$.

Proof. Imagine a string of 0's and 1's, with 1's representing successes.
The prob of any specific string of n 0's and r 1's is $p^r q^n$.

How many such strings are there?

Because we stop as soon as we hit the rth success, the string must terminate in a 1.

Among the other $n + r - 1$ positions, we choose $r - 1$ places for the remaining 1's to go.

So the overall prob of exactly n failures before the rth success is $P(X = n) = \binom{n+r-1}{r-1} p^r q^n$, $n = 0, 1, 2, \ldots$ ■

Ex (**Bike Riders**): Assume 5% of people traveling down a sidewalk are on bikes.

You watch as folks go by. Calculate the prob you'll see 10 non-bikers before you see a third biker?

Solution: $r = 3$ and $p = 0.05$. So: $P(X = 10) = \binom{12}{2}(0.05)^3(1-0.05)^{10} \approx 0.005$ or 0.5%. □

Thm. Let $X \sim NBin(r,p)$ be viewed as # of failures before the rth success in a sequence of indep Bernoulli trials w/success p. We can write $X = X_1 + \ldots + X_r$ where the X_i are iid $Geom(p)$.

[Proof in book]

Expectation of Negative Binomial: $E(X) = E(X_1) + \ldots + E(X_r) = r \cdot \frac{q}{p}$.

What did we learn?

- Expectation of a Discrete Rv
- Geometric Distr/PF/CDF/Expectation
- Negative Binomial Distr/PF/Expectation