Lecture 6:

Expectation

Definition of expectation

Definition of expectation

The **distribution** of a r.v. gives one full information about it. Sometimes we want just one number summarising the “average” value of the r.v. There are several ways to define this “average”, but the most common is the **mean** of an r.v, aka its **expected value**.

Given a list of numbers $x_1, x_2, \ldots, x_n$, their **arithmetic mean** is

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$$

More generally, a **weighted mean** of $x_1, \ldots, x_n$, is $\text{WM}(x) = \sum_{j=1}^{n} x_j p_j$

where the weights $p_1, \ldots, p_n$ are pre-defined nonnegative numbers that add up to 1 (so the unweighted mean $\bar{x}$ is then $p_j = 1/n,\ \forall j$)

Definition of expectation

Definition 4.1.1 (Expectation of a discrete r.v.) The **expected value** (also **expectation** or **mean**) of a discrete r.v. X whose distinct possible values are $x_1, x_2, \ldots$ is:

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j)$$

If the support is finite, this is a finite sum. One can also write

$E(X) = \sum_{x} \underbrace{x}_{\text{value}} \underbrace{P(X = x)}_{\text{PMF at } x}$ where the sum is over the support of X

(in any case, $xP(X = 0)$ for any $x \notin \text{Supp}(X)$).

The expectation is ***undefined*** if $\sum_{j=1}^{\infty} |x_j| P(X = x_j)$ diverges!

Definition of expectation

Examples:

1) X = a fair 6-sided die, so $X = 1,2,3,4,5,6$, with probabilities $1/6$ each. The expected value is $E(X) = \frac{1}{6}(1 + 2 + \ldots + 6) = 3.5$. Note that X ***never*** equals its mean.

2) Let $X \sim$ Bern(p) and $q = 1 - p$. Then $E(X) = 1p + 0q = p$. Imagine a balance (of length) with a pebble of mass q on the left (at 0) and of mass p on the right (at 1). Then for the balance to be stable, the fulcrum should be at p – the **center of mass**.

Frequentist interpretation: we do a series of experiments, writing 1 for “success”, and 0 for “failure” – then p is the proportion of 1-s in our notes.

Definition of expectation

Proposition 4.1.2 If X and Y are discrete r.v.-s with the same distribution, then $E(X) = E(Y)$ (if either side exists).

The converse of this is false – expectation is just a one-number summary, not nearly enough to specify the entire distribution. It's only the ***measure of where the "center" is***.

Do not mix up the random variable, X, and its expected value, $E(X)$.

Notation 4.1.4 We often abbreviate $E(X)$ to EX. Similarly, we often abbreviate $E(X^2)$ to EX^2 and $E(X^n)$ to EX^n.

Mind the ***order*** of operations! EX^2 is the expectation of X^2, not the square of the number EX.

Linearity of expectation

Linearity of expectation

A very important property of expectation is its **linearity**:

Theorem 4.2.1 (Linearity of expectation). For any r.v.-s X and Y and any constant c, $E(X+Y) = E(X) + E(Y)$ and $E(cX) = cE(X)$.

The first equality seems reasonable for independent X and Y, but it holds for any!

One intuition for linearity of expectation is via simulation (sampling): if we sample many times from the distribution of X, the histogram of the sample will look very close to the true PMF of X – in particular, the **arithmetic mean** of sampled values will be very close to the true value of $E(X)$.

Linearity of expectation

Example 4.2.2 (Binomial expectation). For $X \sim$ Bin(n, p). By def,

$E(X) = \sum_{k=0}^{n} kP(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$. Using $kC_n^k = nC_{n-1}^{k-1}$,

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = n \sum_{k=0}^{n} \binom{n-1}{k-1} p^k q^{n-k} =$$

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

Linearity gives another proof: $X = I_1 + \ldots + I_n$ where $I_k \sim$ Bern(p), so $E(X) = E(I_1) + \ldots + E(I_n) = np$.

Linearity of expectation

Example 4.2.2 (Hypergeometric expectation). $X \sim \text{HGeom}(w, b, n)$ = the # of white balls out of n drawn without replacement from urn with w white & b black balls.

As in Binomial, $X = I_1 + \ldots + I_n$, where, by symmetry, $I_k \sim \text{Bern}(p)$ with $p = \dfrac{w}{w+b}$ (unconditionally k-th ball drawn is equally likely to be any of the balls). So, $E(X) = n\dfrac{w}{w+b}$.

Unlike the Binomial, the I_k are ***not independent***, but linearity of expectation still holds for dependent random variables!

Linearity of expectation

Proposition 4.2.4 (Monotonicity of expectation). Let X and Y be such r.v.-s that $X \geq Y$ with probability 1. Then $E(X) \geq E(Y)$, with equality if and only if $X = Y$ with probability 1.

Proof: This holds not only for discrete r.v.-s, by the way.
A r.v. $Z = X - Y$ is non-negative with probability 1, so $E(Z) \geq 0$.
But by linearity,

$$E(X) - E(Y) = E(X - Y) \geq 0.$$

If $E(X) = E(Y)$, then $E(Z) = 0$, then $P(X = Y) = P(Z = 0) = 1$, since if even one term in the sum giving $E(Z)$ is positive, the whole sum is positive.

Geometric and Negative Binomial

Geometric and Negative Binomial

Story 4.3.1 (Geometric distribution). Consider a sequence of independent Bernoulli trials, each with success prob. $p \in (0,1)$, with trials performed until a success occurs. Let X = # of ***failures*** before the first success. Then $X \sim$ Geom(p).

Theorem 4.3.2 (Geometric PMF). If $X \sim$ Geom(p), then the PMF is $P(X = k) = q^k p$ for $k = 0,1,\ldots$, where $q = 1 - p$.

This is a valid PMF, because summing a geometric series gives:

$$p \sum_{k=0}^{\infty} q^k = p \frac{1}{1-p} = 1$$

Geometric and Negative Binomial

Story 4.3.1 (Geometric distribution). Consider a sequence of independent Bernoulli trials, each with success prob. $p \in (0,1)$, with trials performed until a success occurs. Let X = # of ***failures*** before the first success. Then $X \sim \text{Geom}(p)$.

Theorem 4.3.2 (Geometric PMF). If $X \sim \text{Geom}(p)$, then the PMF is $P(X = k) = q^k p$ for $k = 0,1,\ldots$, where $q = 1 - p$.

Theorem 4.3.3 (Geometric CDF). If $X \sim \text{Geom}(p)$, then the CDF is

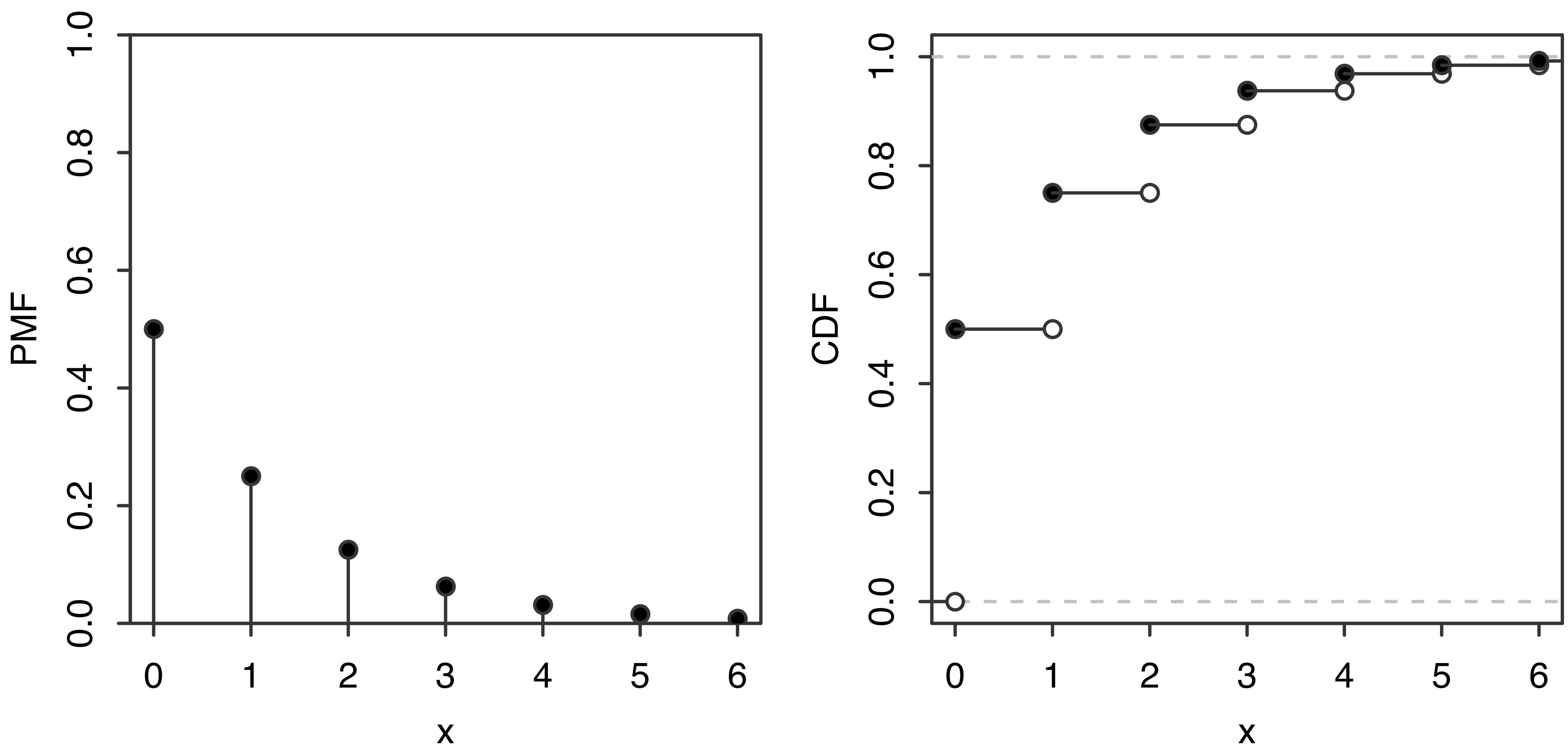
$$F(x) = \begin{cases} 1 - q^{\lfloor x \rfloor + 1}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

where $\lfloor x \rfloor$ = (greatest integer $\leq x$).

Proof is given by summing the PDF.

Geometric and Negative Binomial

Story 4.3.1 (Geometric distribution). Consider a sequence of independent Bernoulli trials, each with success prob. $p \in (0,1)$, with trials performed until a success occurs. Let X = # of ***failures*** before the first success. Then $X \sim$ Geom(p).



Geometric and Negative Binomial

4.3.4 (Conventions for the Geometric). Here, X = # of trials ***before*** the 1st success (not including it). Some authors include it.

Definition 4.3.5 (First Success distribution). In a sequence of independent Bernoulli trials with success prob p, Y = (# of trials until the 1st successful one, including it). Then $Y \sim$ FS(p).

Clearly, if $Y \sim$ FS(p), then $Y - 1 \sim$ Geom(p), and backwards:
if $X \sim$ Geom(p), then $X + 1 \sim$ FS(p).

Geometric and Negative Binomial

Example 4.3.6 (Geometric expectation). Let $X \sim$ Geom(p). We have $E(X) = \sum_{k=0}^{\infty} kq^k p$. Remember that $(q^k)'_q = kq^{k-1}$, so diff-ing $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$ gives $\sum_{k=0}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}$, and so

$$E(X) = \sum_{k=0}^{\infty} kq^k p = pq \sum_{k=0}^{\infty} kq^{k-1} = pq \frac{1}{(1-q)^2} = \frac{q}{p}$$

Example 4.3.7 (First Success expectation). Since for $Y \sim$ FS(p), $Y - 1 \sim$ Geom(p), by linearity, $E(Y) = E(X+1) = \frac{q}{p} + 1 = \frac{1}{p}$

Geometric and Negative Binomial

Story 4.3.8 (Negative Binomial distribution). In a sequence of independent Bernoulli trials with success prob p, X = (# of ***failures*** before the r-th success), then $X \sim$ NBin(r, p).

Theorem 4.3.9 (Negative Binomial PMF). If $X \sim$ NBin(r, p), then

$$P(X = n) = \binom{n + r - 1}{r - 1} p^r q^n$$

Proof: Encode with a string of 0-s and 1-s. Probability of any ***specific*** string = $p^r q^n$. Last symbol should be 1. Among $n + r - 1$ positions, choose $r - 1$ places to put the remaining 1-s to.

Geometric and Negative Binomial

Theorem 4.3.10 Let $X \sim$ NBin(r, p) = (# of failures before r-th success) in a sequence of ind. Bern. trials with success prob. p. Then $X = X_1 + \ldots + X_r$ where X_i are i.i.d. Geom(p).

Example 4.3.11 (Negative Binomial expectation). $X \sim$ NBin(r, p). Using the previous theorem, $X = X_1 + \ldots + X_r$ with $X_i \sim$ Geom(p). Then $E(X) = E(X_1) + \ldots + E(X_r) = r\dfrac{q}{p}$

Example 4.3.12 (Coupon collector). There are n types of toys, you are collecting one by one, the goal is getting a complete set. What's the expected # of toys you need to collect until you complete the set?

Geometric and Negative Binomial

Example 4.3.12 (Coupon collector). There are n types of toys, you are collecting one by one, the goal is getting a complete set. What's the expected # of toys you need to collect until you complete the set?

Let N_1 = time (# of toys) until the 1st new toy type, N_2 = additional time until the 2nd new toy type, etc.

We have $N_2 \sim$ FS$((n-1)/n)$ – after 1st toy type, there's $1/n$ chance of getting the same one. In general $N_j \sim$ FS$((n-j+1)/n)$,

so $E(N) = E(N_1) + \ldots = 1 + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + n = n \sum_{k=1}^{n} \frac{1}{k}$

For large n, this is close to $n(\log n + 0.577)$ (Euler, Mascheroni)

Geometric and Negative Binomial

4.3.13 (Expectation of nonlinear function). Expectation is linear, but in general $E(g(X)) \neq g(E(X))$ for arbitrary $g(x)$.

Example 4.3.14 (St. Petersburg paradox). The game is: you flip a fair coin until 1st Heads. You 2^n dollars if game lasts n rounds. Whats the fair value (expected payoff) of this game?

Solution: Let X = your winnings from playing. So, $X = 2^N$, where N = # of rounds the game lasts. So X = 2 with prob. 1/2, = 4 with prob. 1/4, etc, so $E(X) = \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \ldots = \infty$! On the other hand, the # of rounds N that the game lasts = # of tosses until the 1st Heads, so $N \sim$ FS$(1/2)$ and $E(N) = 2$. Thus $E(2^N) = \infty$ while $2^{E(N)} = 4$

Indicator r.v.s & the fundamental bridge

Indicator r.v.s & the fundamental bridge

Indicator random variable I_A (or $I(A)$) for event A = 1, if A occurs; and 0 otherwise. So $I_A \sim$ Bern$(P(A))$.

Theorem 4.4.1 (Indicator r.v. properties). Let A and B – events. Then

1) $(I_A)^k = I_A$ for any positive integer k
2) $I_{A^c} = 1 - I_A$
3) $I_{A \cap B} = I_A I_B$
4) $I_{A \cup B} = I_A + I_B - I_A I_B$

Proof: For 1), $0^k = 0$ and $1^k = 1$. 2) is from definition. 3) holds because $I_A I_B$ is 1 if both I_A and I_B are 1. For 4), $I_{A \cup B} =$

$$= 1 - I_{A^c \cap B^c} = 1 - I_{A^c} I_{B^c} = 1 - (1 - I_A)(1 - I_B) = I_A + I_B - I_A I_B$$

Indicator r.v.s & the fundamental bridge

Theorem 4.4.2 (Fundamental bridge b/w prob. and expectation). There is a one-to-one correspondence b/w events & ind. r.v.-s, and the prob. of an event A is $P(A) = E(I_A)$ – the expectation of I_A.

Example 4.4.3 (Boole, Bonferroni, inclusion-exclusion). Let $A_1, A_2, \ldots, A_n$ be events. $I(A_1 \cup \ldots \cup A_n) \leq I(A_1) + \ldots + I(A_n)$ – because if LHS = 0 this is true, and if LHS = 1, then at least one term in the RHS = 1. Taking the expectation,

$$P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + \ldots + P(A_n)$$

– this is called Boole's or Bonferroni's inequality.

To prove inclusion-exclusion, we can use properties of indicator random variables on $1 - I(A_1 \cup \ldots \cup A_n)$

Indicator r.v.s & the fundamental bridge

Example 4.4.4 (Matching continued). A well-shuffled deck of n cards, labeled 1 through n. A card is a ***match,*** if its position = its label. Let X = # of matches, find $E(X)$.

Solution: Actually, X follows none of the distributions we studied so far, but $X = I_1 + \ldots + I_n$, where

$$I_k = \begin{cases} 1, & \text{if } k\text{-th card is a match} \\ 0, & \text{otherwise} \end{cases}$$

So $E(I_k) = P(j\text{-th card - match}) = \dfrac{1}{n}$ and so

$E(X) = E(I_1) + \ldots + E(I_n) = n \cdot \frac{1}{n} = 1$ – doesn't depend on n!

Law of the unconscious statistician

(LOTUS)

Law of the unconscious statistician (LOTUS)

As we saw, in general $E(g(X)) \neq g(E(X))$ (if g is non-linear). But it turns, we ***don't need*** the PMF/CDF of $g(X)$ to compute $E(g(X))$

Theorem 4.5.1 (LOTUS). If X is a discrete r.v., g is a function $g : \mathbb{R} \to \mathbb{R}$. Then $E(g(X)) = \sum_{x \in \text{Supp}(X)} g(x)\, P(X = x)$

The name is so b.c. we just replace $x \to g(x)$ in the sum

Even if g is not a one-to-one function, this works – we sum over the support of x, not of $g(x)$