Lecture 2: Expectations & Probability Review.

Subject :

Date:

X: non-negative discrete rv. Sample values are $\{a_1, a_2, \ldots\} \subset \mathbb{R}_+$

$$\mathbb{E}[X] = \sum_{j=1}^{\infty} a_j p_X(a_j) \quad \text{where } p_X(\cdot) \text{ is the pmf of } X.$$

Expectation is said to exist if $\mathbb{E}X < \infty$.

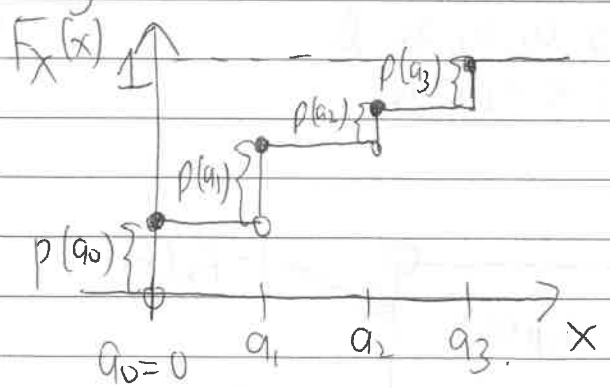
Expectation need not exist if # of sample values is infinite (see book)

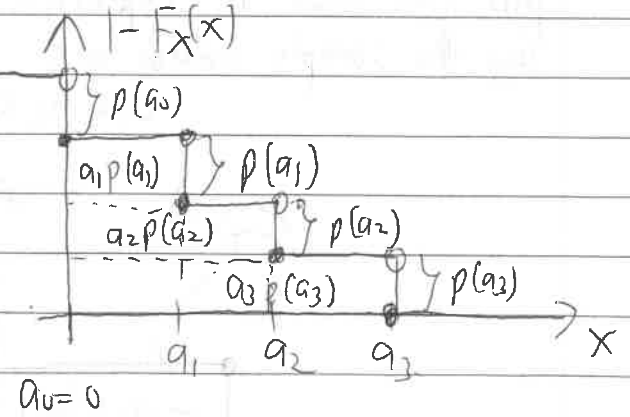
Eg: $\Pr(X=n) = \frac{1}{n(n+1)}$, $n \in \mathbb{N}$. Check that $\sum_{n\in\mathbb{N}} \Pr(X=n) = 1$.

$$\mathbb{E}X = \sum_{n\in\mathbb{N}} n\, p_X(n) = \sum_{n\in\mathbb{N}} \frac{1}{n+1} = \infty.$$

Fact: $\mathbb{E}X = \int_0^\infty 1 - F_X(x)\,dx$ where $F_X(x) = \Pr(X \le x)$ is the cdf.

Pf: What does the cdf look like for a non-negative discrete rv with finitely many sample pts? Say $\{a_0 = 0, a_1, a_2, a_3\} \subseteq \mathbb{R}_+$.
Say $a_0 = 0 < a_1 < a_2 < a_3$.





Area under the graph of $x \mapsto 1 - F_X(x)$ is

$$a_1 p(a_1) + a_2 p(a_2) + a_3 p(a_3) = \int_0^\infty 1 - F_X(x)\,dx.$$

Fact: For a non-negative integer-valued rv X,

$$\mathbb{E}X = \sum_{n=0}^{\infty} \Pr(X > n) = \sum_{n=1}^{\infty} \Pr(X \ge n)$$

Pf: $\sum_{n=0}^{\infty} P_r(X>n) = P_r(X>0) + P_r(X>1) + P_r(X>2) + \cdots$

$= P_r(X \geq 1) + P_r(X \geq 2) + P_r(X \geq 3) + \cdots$

$= (p(1)+p(2)+p(3)+\cdots) + (p(2)+p(3)+\cdots) + (p(3)+\cdots)$

$= 1\cdot p(1) + 2\cdot p(2) + 3\cdot p(3) + 4\cdot p(4) + \cdots$

$= \sum_{n=1}^{\infty} n\, p_X(n) = \mathbb{E}X.$

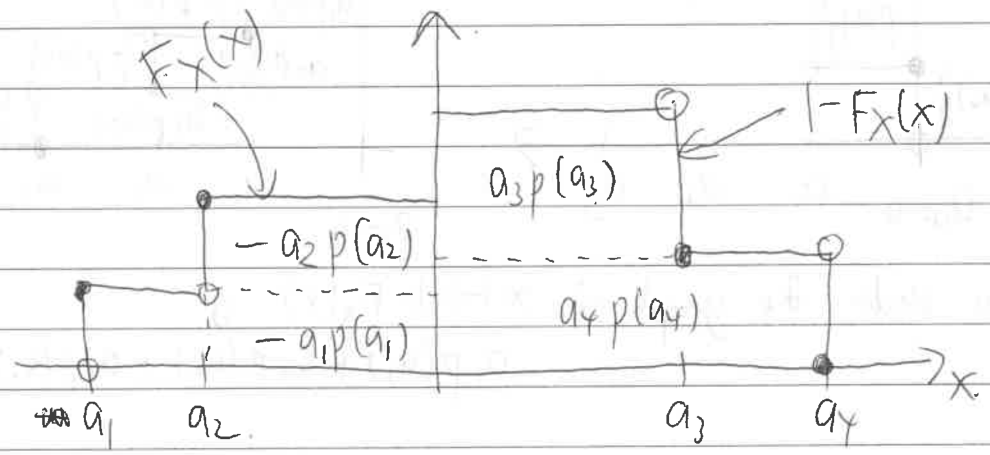
Thus we may alternatively write the expectation of a non-negative rv as

$$\mathbb{E}X = \int_0^{\infty} F_X^c(x)\,dx = \int_0^{\infty} 1 - F_X(x)\,dx, \quad F_X^c(x) = P_r(X>x)$$

What if X has both positive & negative sample values?
$X \in \{a_1, a_2, \ldots, a_n\} \subset \mathbb{R}$.

$$\mathbb{E}X = \sum_i a_i\, p(a_i) = \sum_{i;\,a_i \leq 0} a_i\, p_X(a_i) + \sum_{i;\,a_i>0} a_i\, p_X(a_i).$$

This can also be expressed in terms of the cdf of X, $F_X(x)$.
Say the sample values are a_1, a_2, a_3, a_4 &
$a_1 < a_2 < 0 < a_3 < a_4$.



$$\mathbb{E}X = \int_{-\infty}^{0} -F_X(x)\,dx + \int_0^{\infty} 1 - F_X(x)\,dx \qquad u(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x<0 \end{cases}$$

$$= \int_{-\infty}^{\infty} \left(u(x) - F_X(x)\right) dx \quad -(*)$$

Subject: X

Date:

Def: The expectation of a rv $\mathbb{E}X$ exists, with the value given in (*) if each of the above 2 terms is finite.

Random variables as functions of other rvs

X: rv $\quad h: \mathbb{R} \to \mathbb{R}$.

Define $Y = h(X)$. We can evaluate the expectation of Y in 2 ways.

i) $\mathbb{E}Y = \int h(x) f_X(x)\, dx$, $\quad$ ii) $\mathbb{E}Y = \int y \underbrace{f_Y(y)}_{\text{pdf of } Y}\, dy$

Some functions are more impt than others.

$h(x) = x^n$ yield the "moments" of X; $h(x) = (x - \bar{x})^2$, $\bar{x} = \mathbb{E}X$ (gives the variance)

$\mathbb{E}Y = \mathbb{E}h(X) = \mathbb{E}[(X - \bar{x})^2] = Var(X) = \sigma^2$

$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

Important connection btw. mean & variance

$\min_{x \in \mathbb{R}} \mathbb{E}[(X-x)^2]$ is achieved at $x^* = \mathbb{E}X$.

Another function of interest is $h(x,y) = x+y \Rightarrow Z = X + Y$.

If X & Y are independent, we can express the distribution of Z in terms of those of X & Y.

$F_Z(z) = P(Z \le z) = P(X+Y \le z) = \int f_X(x) P(X+Y \le z \mid X=x)\, dx$. (law of total prob).

$= \int f_X(x) P(x + Y \le z \mid X = x)\, dx = \int f_X(x) P(Y \le z-x)\, dx$.

X is fixed to x in the integral $\qquad Y \perp\!\!\!\perp X$

Subject : Date:

$$= \int f_X(x) F_Y(z-x)\, dx.$$

Differentiate both sides wrt z

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \int f_X(x) F_Y(z-x)\, dx$$

$$= \int f_X(x) F_Y'(z-x)\, dx$$

$$= \int f_X(x) f_Y(z-x)\, dx$$

Thus the pdf of Z is the convolution of the pdfs of X & Y. (independence of X & Y is crucial here!)

Important relations: If $X_1, \cdots, X_n$ are independent (in fact uncorrelatedness is enough),

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i).$$

Regardless of the independency or dependency of $X_1, \cdots, X_n$, we have

$$\mathbb{E}[X_1 + \cdots + X_n] = \sum_{i=1}^{n} \mathbb{E} X_i \quad \text{(linearity of expectation).}$$

Example of the ~~util~~ utility of the linearity of expectation.

Def: A graph (V, E) consists of a set $V = \{1, 2, \cdots, n\}$ of vertices and a set of pairs of vertices (called edges) $E \subset \binom{V}{2}$

Def: A hypergraph $H = (V, E)$ consists of a set $V = \{1, \cdots, n\}$ of vertices and a set of tuples of nodes (vertices) E. Each $e \in E$ is known as a hyperedge.

Eg: If $n = 10$, $V = \{1, 2, \cdots, 10\}$.
$E = \{\{1, 2, 3\}, \{4, 5\}, \{4, 5, 6\}, \{3, 4, 7\} \{5, 6, 7, 8, 9\}, \{9, 10\}, \{10\}\}.$

Subject: Date:

Def: A hypergraph H is monochromatic if all the nodes have the same color.

Def: A hypergraph H is called ~~k-regular~~ k-uniform if each edge $e \in E$ contains precisely k nodes.
Eg: $E = \{\{1,2,3\}, \{3,4,5\}, \{4,5,7\}, \{8,9,10\}\}$.

Def: H is 2-colorable if we can color each of the nodes in V so that no edge is monochromatic.

Rmk: If H has more edges, it is "less likely" to be 2-colorable.

Thm: Every k-uniform (k-regular) hypergraph with $< 2^{k-1}$ edges is 2-colorable.

Rmk: Erdős proved this in 1963.
Erdős also showed that $\exists$ a ~~k-regular~~ k-uniform hypergraph with $O(2^k k^2)$ hyperedges that is not 2-colorable. So the base 2 is sharp.

Pf: Let $H=(V,E)$ be a k-uniform hypergraph with $< 2^{k-1}$ edges.

Color each node $v \in V$ with one of 2 colors red / blue with equal probability. For each hyperedge $e \in E$, define the r.v.

$$X_e = \begin{cases} 1 & e \text{ is monochromatic} \\ 0 & \text{else} \end{cases}$$

Consider $X = \sum_{e \in E} X_e$, the total # of monochromatic edges in H.

$\mathbb{E}X_e = 1 \cdot \Pr(e \text{ is monochromatic}) + 0 \cdot \Pr(e \text{ is not monochromatic})$
$= 1 \cdot 2 \cdot 2^{-k} = 2^{-k+1}$

Subject : Date:

Hence $\mathbb{E}X =$ (num of edges in H) $\cdot 2^{-k+1} < 2^{+k-1} \cdot 2^{-k+1} = 1$

Note that $\mathbb{E}X = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) + \cdots$

Since $\mathbb{E}X < 1$, $\exists$ $p_X(0) > 0$. This means that the prob. that H has 0 monochromatic edges (i.e, that it is 2-colorable) is positive.

Hence $\exists$ a coloring of H with no monochromatic edges, meaning H is 2-colorable.

Conditional Expectation.

Say X, Y are discrete r.v's with $p_Y(y) > 0$ $\forall y$.

Conditional Expectation of X given $Y=y$ is

$$\mathbb{E}[X|Y=y] = \sum_x x \, p_{X|Y}(x|y)$$

Conditional cdf of X given $Y=y$ (for y s.t. $p_Y(y)>0$)

$$F_{X|Y}(x|y) = \Pr(X \leq x, Y=y) / \Pr(Y=y).$$

What's the meaning of the conditional expectation of X given a rv Y?

Define $g(y) = \mathbb{E}[X|Y=y]$. This is a f^n of y.

Thus $\mathbb{E}[X|Y]$ is the rv. $g(Y)$

Subject : Date:

Fact (Law of iterated expectation)

$$\mathbb{E}_Y[\mathbb{E}_X[X|Y]] = \mathbb{E}_X[X]$$

Pf: $\text{LHS} = \sum_y p_Y(y)\,\mathbb{E}[X|Y=y] = \sum_y p_Y(y) \sum_x x\, p_{X|Y}(x|y)$.

$$= \sum_x x \sum_y p_Y(y)\, p_{X|Y}(x|y) = \sum_x x \sum_y p_{XY}(x,y).$$

$$= \sum_x x\, p_X(x) = \mathbb{E}X.$$

Ex: X_1: # of top face of 1^{st} dice. $S := X_1 + X_2$
X_2: # —"— 2^{nd} dice.

Conditioned on $S = j \in \{1,2,\dots,7\}$, $X_1 \sim \text{Unif}\{1,\dots,j-1\}$.
Conditioned on $S = j \in \{8,9,\dots,12\}$, $X_1 \sim \text{Unif}\{j-6,\dots,6\}$.

$$\mathbb{E}[X_1|S=j] = \begin{cases} j/2 & j \le 7 \\ j/2 & j > 7 \end{cases} \Rightarrow \mathbb{E}[X_1|S=j] = j/2.$$

Thus $\underbrace{\mathbb{E}[X_1|S]}_{=Y}$ is a discrete rv (a function of the rv S) taking values 1 to 6 in steps of $1/2$ as the sample values of S go from 2:12..

$p_Y(j/2) = p_S(j)$, $j = 2,3,\dots,12$.

~~$p_Y(y)$~~
$p_Y(y) > 0$
~~$y = 1:0.5:6$~~
$\forall\ y = 1:0.5:6$.

$$\mathbb{E}[X_1] = \mathbb{E}_S[\mathbb{E}[X_1|S]] = \sum_{j=2}^{12} j/2 \cdot p_Y(j/2)$$

$$= \sum_{j=2}^{12} j/2\, p_S(j) = \frac{1}{2}\mathbb{E}S = 7/2.$$

Subject : Date:

Moment Generating Function (MGF).

Given
~~Given~~ X, a rv, we may define $g_X(r) = \mathbb{E}[e^{rX}] = \int_{-\infty}^{\infty} e^{rx} f_X(x)\, dx$.

where $f_X(x)$ is the pdf of X (assumed continuous).

Discrete rv X $\Rightarrow$ $g_X(r) = \sum_x e^{rx} p_X(x)$.

Cumulant generating function is $\ln g_X(r)$.

If $g_X(r)$ exists in a neighborhood around 0, then its derivatives of all orders exist in that nbd.

$$\frac{d^k g_X(r)}{dr^k} = \int_{-\infty}^{\infty} x^k e^{rx} f_X(x)\, dx$$

$$\& \quad \left.\frac{d^k g_X(r)}{dr^k}\right|_{r=0} = \int_{-\infty}^{\infty} x^k f_X(x)\, dx = \mathbb{E}[X^k]$$

Thus we can obtain all moments $\mathbb{E}[X^k]$ (indexed by $k \in \mathbb{N}$) given the MGF.

What's the MGF of $S_n = X_1 + \cdots + X_n$ (X_i's mutually indep)

$$g_{S_n}(r) = \mathbb{E}[e^{rS_n}] = \mathbb{E}[e^{r(X_1 + \cdots + X_n)}]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n} e^{rX_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{rX_i}] = \prod_{i=1}^{n} g_{X_i}(r)$$

If X_i's are i.i.d, then

$$g_{S_n}(r) = (g_X(r))^n.$$

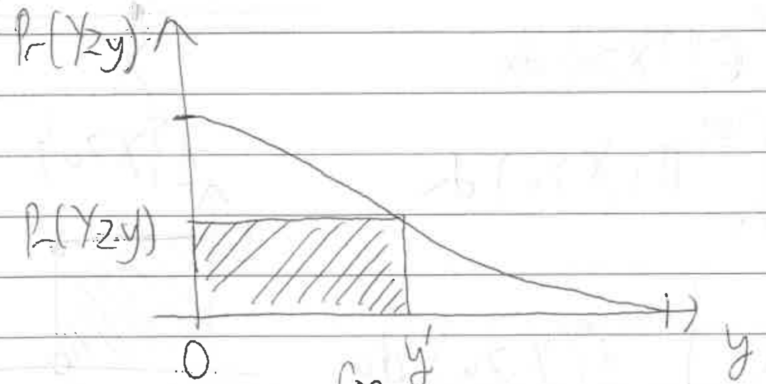
Subject : Date:

Basic Inequalities: Markov's inequality.

Y: non-negative rv $\mathbb{E}Y$: expectation of Y.

Claim: $\Pr(Y \geq y) \leq \frac{\mathbb{E}Y}{y} \quad \forall y > 0.$

Pf:



Recall $\mathbb{E}Y = \int_0^\infty \Pr(Y \geq y')\, dy' \geq y' \Pr(Y \geq y')$

$$\Rightarrow \Pr(Y \geq y') \leq \frac{\mathbb{E}Y}{y'}. \quad \forall y' > 0.$$

Alternatively, consider

$$\mathbb{1}\left\{\frac{Y}{y} \geq 1\right\} \leq \frac{Y}{y} \quad \text{a.s.} \quad -(*)$$

Justification: Consider two cases $Y \geq y$ & $Y < y$

Now ~~take~~ take expectation on both sides

$$\mathbb{E}\left[\mathbb{1}\left\{\frac{Y}{y} \geq 1\right\}\right] \leq \mathbb{E}\left[\frac{Y}{y}\right]$$

$$\Rightarrow \Pr(Y \geq y) \leq \frac{1}{y}\mathbb{E}Y.$$