ASEN 5044, Fall 2024

Statistical Estimation for Dynamical Systems

Lecture 10: Random Variables, Probability Distributions, Density Functions, Expected Values

Prof. Nisar Ahmed (Nisar.Ahmed@Colorado.edu)

Thurs 09/26/2024

Ann and H.J. Smead
Aerospace Engineering Sciences
UNIVERSITY OF COLORADO BOULDER

Announcements

- **HW 3 Due Fri 09/27 (tomorrow)**
- **HW 4: Out today, Due Thurs 10/03 [1 week]**
- **Quiz 4: out tomorrow Fri 09/27, due Tues 10/01**
- **First advanced topic lecture: tomorrow Fri 9/27 (pre-recorded, to be posted)**
  - ***Optional:*** **focused on Bayesian estimation and related topics**
  - **Will post blank + written slides**
- **Midterm 1: out next Thurs 10/03, due Thurs 10/10 (Gradescope)**
  - One week long take home exam posted to Canvas
  - Open book/notes – honor code applies (must complete by yourself)
  - Will cover HW 1-4 (solns will be posted) + Quizzes 1-4 + associated lectures
  - No office hours 10/07-10/10 for TFs or Prof. Ahmed (send private email/Piazza posts for clarification questions only)
  - Expected effort: ~5.5 - 8 hrs

Last Time…

- Joint probabilities
- Marginal probabilities
- Conditional probabilities
- Bayes' rule
- Independent/dependent ~~random variables~~ events

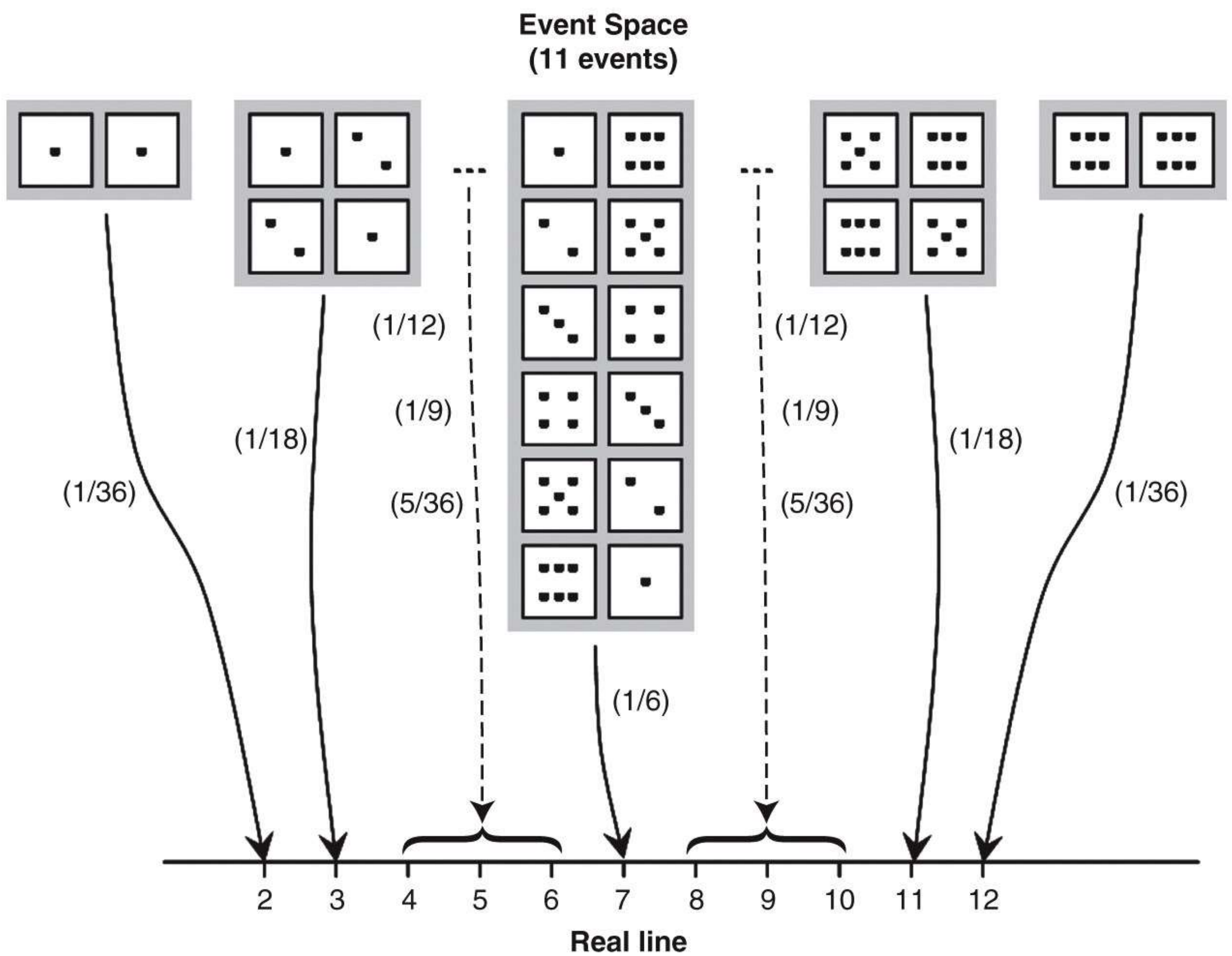
Today…

- discrete and continuous random variables (i.e. “random quantities”)
- probability mass functions (pmfs) for discrete random variables
- probability density functions (pdfs) for continuous random variables
- Expectation operators and expected values, examples

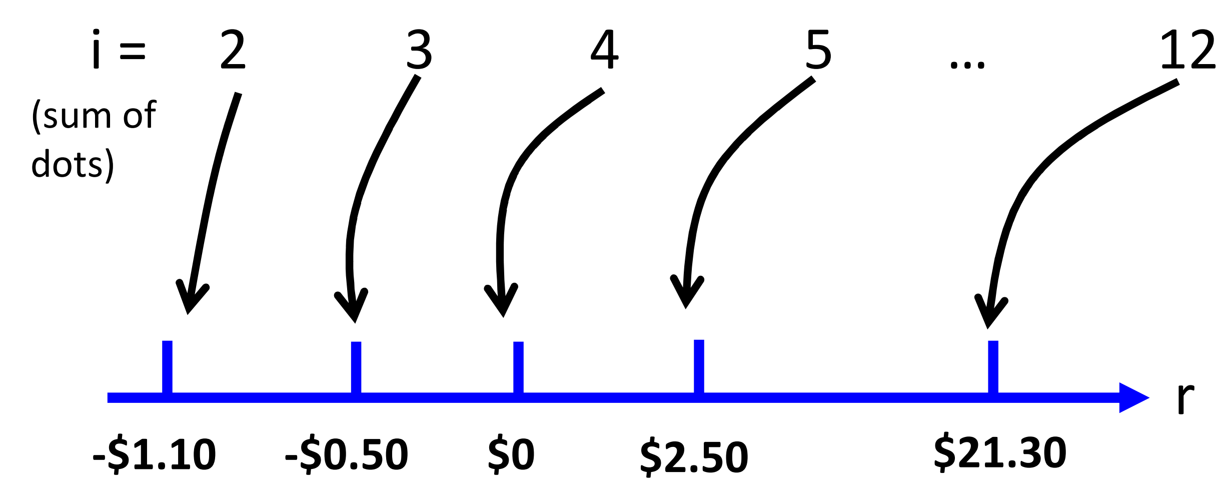
READ SIMON BOOK, CHAPTER 2.5

Random Variables

- A **random variable (RV)** is a function that maps every point in an event space $\{A_i\}$ to points on the real line
- Example: RVs for two dice: Fxn #1: add up dots



Fxn #2: Suppose you receive arbitrary reward R for getting certain # of dots
→ can assign R=r to be a random variable



What's the Point of Defining RVs?

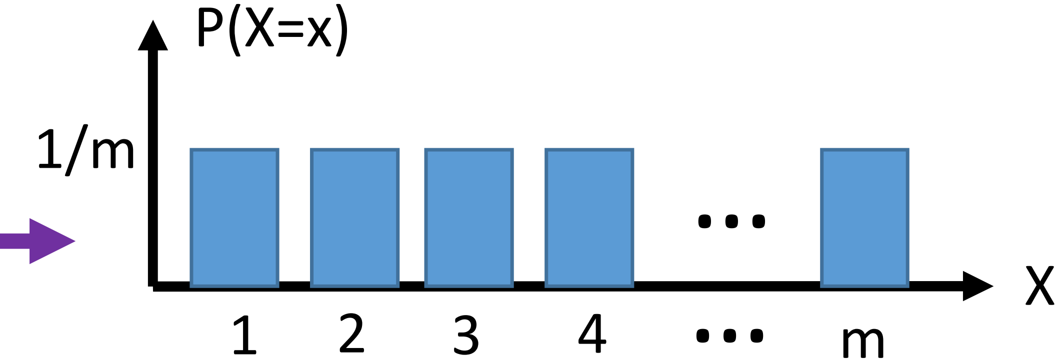
- Much easier to work with/visualize probabilities on RVs than "raw" events and outcomes
- Think of **"random quantity"** as another name for a "random variable"
- Other examples of random quantities (or RVs) that can be readily assigned to otherwise non-quantitative outcomes/events for random experiments:
  - Select a person on the street at random & then **measure** their height **h**, or weight **w**, or age **a**, or GPA **g**,…
    → any particular person is now "quantified" by a number on the real line
    → example of **continuous random quantity (i.e. a continuous RV)**
  - Flip a coin 5 times & then count (**i.e. measure**) number of heads → any particular outcome (e.g. THHHH, HTHHH, HHTHH,…) now maps to a number on the real line (integer in this case)
    → example of **discrete random quantity (i.e. a discrete RV)**
  - Take a reading from a Geiger counter and report the value you see on the dial
    → continuous RV (identity mapping)

Discrete Random Variables

X is a discrete RV if X maps outcomes/events to integer quantities

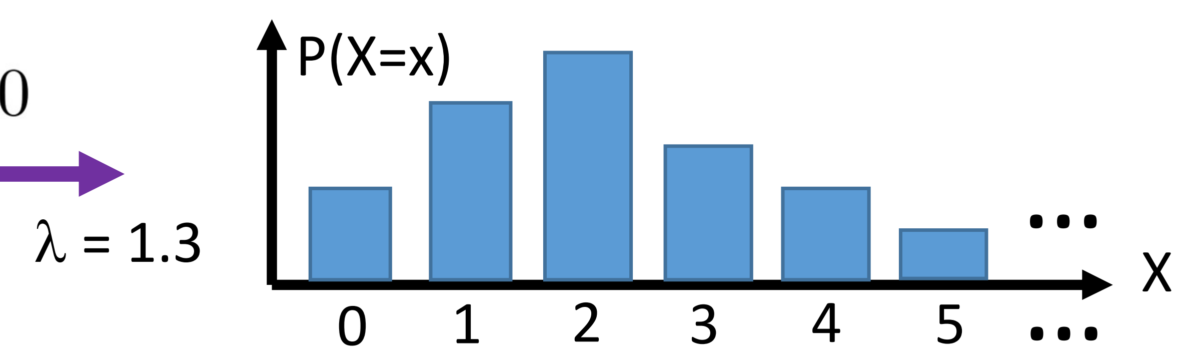
- Can be **finite or countably infinite** (e.g. number of e-mails between now and midnight)
- A fxn that assigns a single probability to each possible realization x of X, i.e. P(X=x), is called a **probability mass function (pmf)**
- The **pmf** is also sometimes called a **discrete probability distribution**
- **Example discrete probability distributions (pmfs):**

Uniform: $P(X=x) = \frac{1}{m},\ \text{for } x \in \{1, ..., m\}$



Poisson: $P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!},\ \text{for } x \in \{0, 1, 2, ..., \},\ \lambda \geq 0$

(→countably infinite outcomes, e.g. # arrivals per hour)



Bernoulli: $P(X=x) = p^x(1-p)^{1-x},\ \text{for } x \in \{0, 1\}$

(→binary outcomes, e.g. coin flips; p = probability of x=1)

Binomial: $P(X=x) = \frac{n!}{x!(n-x)!} p^x(1-p)^{n-x},\ \text{for } x \in \{0, 1, ..., n\}$ (→probability on total # of "1's" in a sequence of n Bernoulli trials)

Continuous Random Variables

X is a continuous RV if it maps to continuous quantities (real-valued, for our purposes)

- **Uncountably infinite** (e.g. there is a continuum of numbers between 100 and 100.1)
- We need to be careful about assigning and defining what P(X=x) really means!!!
- Example: spinning the pointer on a wheel

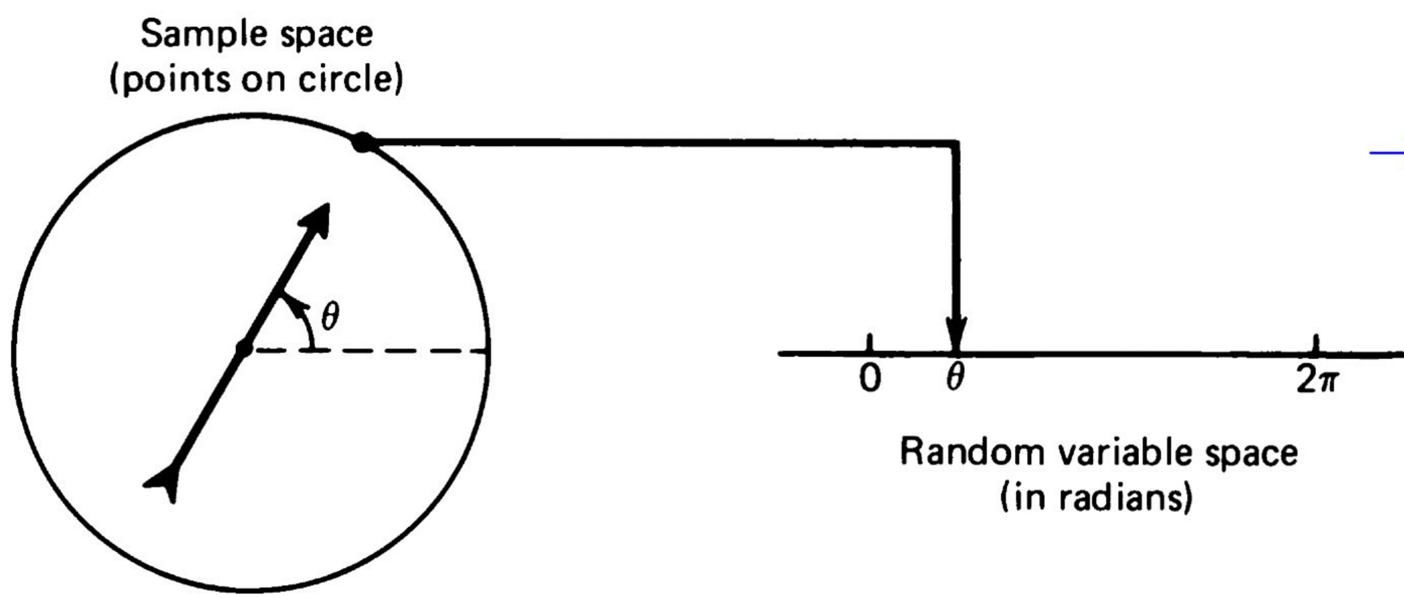


Figure 1.3a

All possible θ are likely [if wheel is fair]

$\rightarrow$Naive '$\frac{1}{N}$' relative frequency calculation of prob:

$$\text{Prob}(\theta) = \lim_{N\rightarrow\infty} \frac{1}{N} = 0$$

$\rightarrow$But then sum over $\text{Prob}(\theta)$ for all $\theta = 0$

$\rightarrow$Violates axioms of probability:

need sum of $\text{Prob}(\theta)$ for all θ to be 1!!

- **Recall:** probabilities defined on **events** for outcome space -- so we **need a way to properly define events over a continuum of outcomes,** and then assign probabilities to such events …
- Most natural way: define events to be **intervals (lengths) on continuous real line**
- So then we **need a way to assign probabilities to arbitrary intervals (events) on real line**

Probability Density Function (pdf)

- A fxn that assigns a single probability to each possible interval (x_1,x_2) of X=x, i.e. $P(x_1 < x < x_2)$, is called a **probability density function (pdf)**
  - The **pdf** is also sometimes called a **continuous probability distribution**
  - **Since probability is dimensionless, it follows that the pdf must have units = 1/[units of X]**
  - **Example pdfs:** uniform, Gaussian, exponential, Gamma, Beta, Rayleigh, Student's-t, Laplace, Weibull…
- Formally:
  - Event: $\{x : \xi - d\xi < x < \xi\}$ (i.e. either x falls inside the interval $[\xi - d\xi, \xi]$, or it does not)
  - The **probability density function (pdf)** of a scalar RV: $\triangleq$ (defined as)

$$\lim_{d\xi \to 0} \frac{P(\xi - d\xi < x < \xi)}{d\xi} \equiv p_x(\xi) = p_X(x) = p(x)$$

  - From axioms of probability, it follows that: $P(\eta < x < \xi) = \int_\eta^\xi p(x)dx \; = c(\xi) - c(\eta)$
  - **Cumulative distribution function (cdf)**: $P(-\infty < x \leq \xi) = \int_{-\infty}^{\xi} p(x)dx \equiv c(\xi)$

$$p(x) = \frac{dP(-\infty < x \leq \xi)}{dx}$$

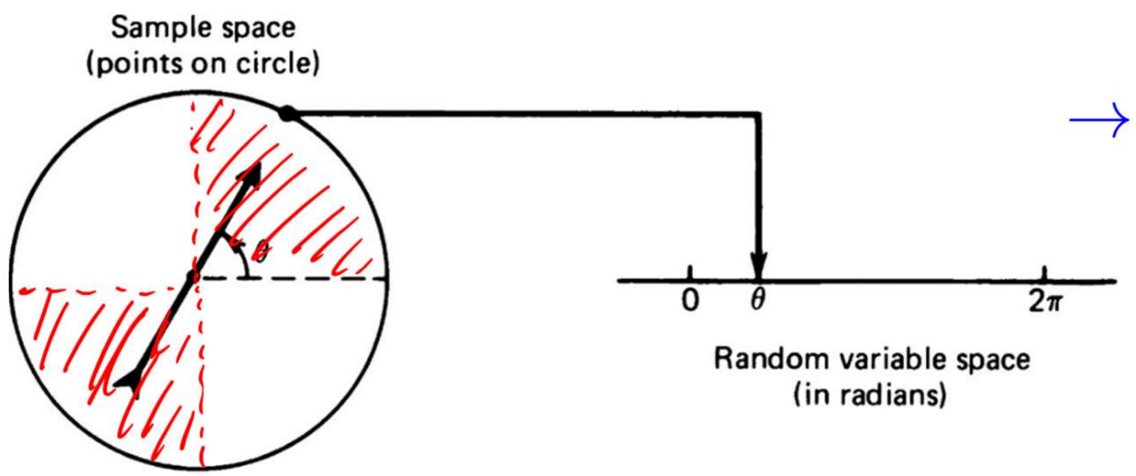
(i.e. pdf is derivative of cdf, when cdf is continuous and differentiable)

  - From probability axioms, must always have: $\int_{-\infty}^{\infty} p(x)dx = 1$

PDF Example: Spinning Pointer on Wheel

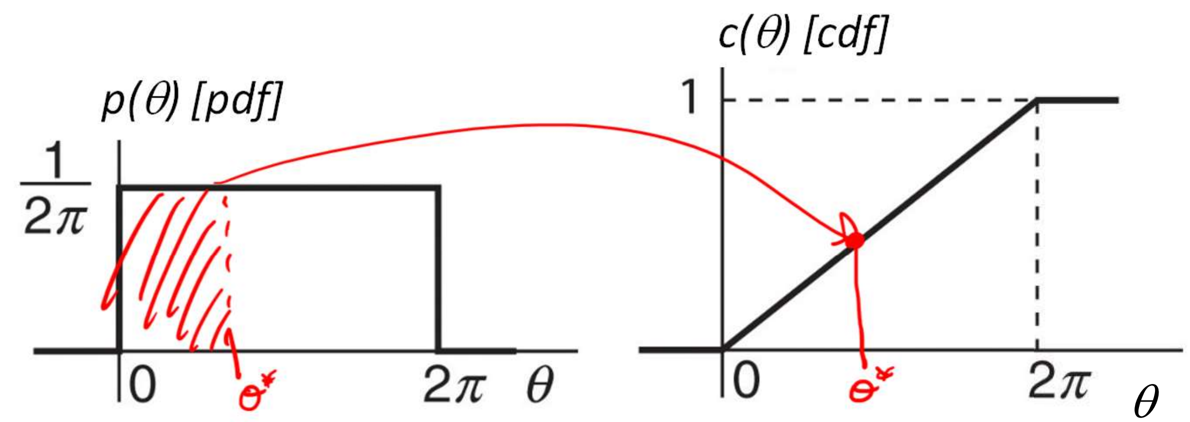
"distributed as"

- If spinner fairly constructed, then θ has uniform pdf: $\theta \sim \mathrm{U}[a, b]$ ('$\sim$': 'distributed as'; $a = 0$, $b = 2\pi$)



$$\rightarrow p(\theta) = \mathrm{U}_\theta[0, 2\pi]$$

$$= \begin{cases} \frac{1}{2\pi}, & \theta \in [0, 2\pi] \\ 0, & \text{o'wise} \end{cases}$$



(i) What is $P(0 \leq \theta \leq \frac{\pi}{2}) = \ ?$

$$P(0 \leq \theta \leq \frac{\pi}{2}) = \int_0^{\frac{\pi}{2}} p(\theta) d\theta = \int_0^{\frac{\pi}{2}} \frac{1}{2\pi} d\theta = \frac{1}{2\pi} \int_0^{\frac{\pi}{2}} 1 \ d\theta = \frac{1}{2\pi} \cdot \theta|_0^{\pi/2} = \frac{1}{4}$$

(ii) What is $P([0 \leq \theta \leq \frac{\pi}{2}] \text{ OR } [\pi \leq \theta \leq \frac{3\pi}{2}]) = \ ?$

Because intervals (events) are disjoint:

$$P(0 \leq \theta \leq \frac{\pi}{2}) + P(\pi \leq \theta \leq \frac{3\pi}{2}) \quad (\text{where } P(\pi \leq \theta \leq \frac{3\pi}{2}) = \int_\pi^{\frac{3\pi}{2}} \frac{1}{2\pi} d\theta = \frac{1}{2\pi} \cdot \theta|_\pi^{3\pi/2} = \frac{1}{4})$$

$$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Expected Values and the Expectation Operator

- Not surprisingly, the function Y = g(X) of a RV X is also a RV (i.e. Y is a RV)
- We could try to find the distribution of Y, but this can be difficult or unnecessary
- Sometimes we just need a "summary of what to expect" from Y without enumerating all possible values for Y
- i.e. what is the "average value" of some arbitrary function g(x) of random var X?

Discrete Case

$$E[g(x)] = \sum_{i=1}^{N_x} g(x=i)P(x=i)$$

= single constant w.r.t. x $\in \mathbb{R}$
(where N_x is the # possible outcomes for x)

Continuous Case

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)p(x)dx$$

= single constant w.r.t. x $\in \mathbb{R}$

$E[\cdot]$ = expectation operator: returns the expected value of g(x) [where $\cdot = g(x)$ i.e. argument of E]
($\mathbb{E}$)

Interpretation of Expected Values

- Consider the "relative frequency" view of probabilities (for discrete RVs):

$$p_i = \lim_{N\to\infty} \frac{N_i}{N}, \qquad x_i \in \{1,2,3,\cdots,N_x\} \Rightarrow \# \text{ times we } \textit{typically} \text{ expect to see } x_i \text{ in } N \to \infty \text{ trials:}$$

(prob. of RV x_i for event $i = 1,\cdots,N$, given N_i occurrences in N trials)

$$N_1 = p_1 \cdot N \text{ (typical \# times } x_i = 1)$$
$$N_2 = p_2 \cdot N \text{ (typical \# times } x_i = 2)$$
$$\vdots$$
$$N_{N_x} = p_{N_x} \cdot N \text{ (typical \# times } x_i = N_x)$$

- So given some sample of outcomes, the 'typical' ***N-sample mean*** for corresponding RVs would be:

$$\bar{x}_{sample} \equiv \frac{(N_1 \cdot x_1) + (N_2 \cdot x_2) + \cdots + (N_{N_x} \cdot x_{N_x})}{N} \underset{(\text{as } N\to\infty)}{\Rightarrow} \frac{([p_1 \cdot N]\cdot x_1) + ([p_2 \cdot N]\cdot x_2) + \cdots + ([p_{N_x} \cdot N]\cdot x_{N_x})}{N}$$

(typical sample mean, i.e. typical sample avg. value for arbitrary finite sample size N)

$$= (p_1 \cdot x_1) + (p_2 \cdot x_2) + \cdots + (p_{N_x} \cdot x_{N_x})$$
$$= \sum_{i=1}^{N_x} x_i \cdot p_i = E[x_i]$$

(no dependence on N!!)

- The **expected value (EV)** = *conceptual average* obtained over *infinite # of trials* N
  - Key idea: don't actually need to run infinite # of trials N if we know probability of outcomes
  - EV is what you expect to see in a "typical" random trial (not what you actually will see – b/c trial is random!)
  - Expected value is NOT the same as the sample average (sample mean) for finite N
  - Expected value says NOTHING about the actual number you will obtain for finite N

Some Common/Important Expected Values

Things you will see a lot in estimation problems:

- **1st moment, aka the mean or average of RV x:** $E[x] = \bar{x} = \mu_x = \begin{cases} \int_{-\infty}^{\infty} x p(x) dx & \text{(cont RV)} \\ \sum_{i=1}^{N_x} x P(x) & \text{(disc RV)} \end{cases}$ $(g(x) = x)$

- **2nd moment:** $E[x^2] = \begin{cases} \int_{-\infty}^{\infty} x^2 p(x) dx & \text{(cont RV)} \\ \sum_{i=1}^{N_x} x^2 P(x) & \text{(disc RV)} \end{cases}$
  $(g(x) = x^2)$

- **Variance (aka 2nd moment about the mean):** $\text{var}(x) = \sigma_x^2 \overset{*}{\equiv} E[(\underbrace{x - \mu_x)^2}_{g(x)}] = \begin{cases} \int_{-\infty}^{\infty} (x - \mu_x)^2 p(x) dx & \text{(cont RV)} \\ \sum_{i=1}^{N_x} (x - \mu_x)^2 P(x) & \text{(disc RV)} \end{cases}$
  $(\text{var}(x) \geq 0)$

- **Standard deviation:** $\text{std}(x) = \sigma_x = \sigma = \sqrt{\text{var}(x)} \geq 0$

- **Higher order nth moment** (pdf/pmf shape info, e.g. skewness, kurtosis): $(g(x) = x^n)$ $E[x^n] = \begin{cases} \int_{-\infty}^{\infty} x^n p(x) dx & \text{(cont RV)} \\ \sum_{i=1}^{N_x} x^n P(x) & \text{(disc RV)} \end{cases}$

- **Expected cost/reward fxn value J(x):** $E[J(x)] = \begin{cases} \int_{-\infty}^{\infty} J(x) p(x) dx & \text{(cont RV)} \\ \sum_{i=1}^{N_x} J(x) P(x) & \text{(disc RV)} \end{cases}$

Useful Properties of Expectations (both continuous and discrete)

- FACT 1: The expectation operator is linear

$$E_x\left[\alpha f(x) + \beta g(x)\right] = \alpha \cdot E_x[f(x)] + \beta \cdot E_x[g(x)]$$

for any constants α, β and (integrable/summable) fxns $f(x), g(x)$

- FACT 2: Variance can always be computed more simply as

$$\begin{aligned} \text{var}(x) = E_x\left[(x-\mu_x)^2\right] &= E[x^2] - (E[x])^2 \\ &= E[x^2] - (\mu_x)^2 \end{aligned}$$

Example #1: Die Rolls

(a) Compute the expected face value i for the roll of a single fair die

$$\bar{x} = \mu_x = E[x] = \sum_{i=1}^{6} x_i \cdot P(x_i) \quad = 1 \cdot P(x_i = 1) + 2 \cdot P(x_i = 2) + \cdots + 6 \cdot P(x_i = 6)$$

$$= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{1}{6} \cdot 21$$

$$\boxed{\rightarrow \bar{x} = \mu_x = E[x] = 3.5}$$

(b) Find expected reward (“expected take”) for a single roll, if given reward function R(i)

i	R(i), $
1	0
2	5
3	10
4	10
5	5
6	0

$$\rightarrow E[R] = \sum_{i=1}^{6} R(i) \cdot P(i)$$

$$= R(i=1) \cdot P(i=1) + R(i=2) \cdot P(i=2) + \cdots + R(i=6) \cdot P(i=6)$$

$$= 0 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 10 \cdot \frac{1}{6} + 10 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 0 \cdot \frac{1}{6}$$

$$= (2) \cdot 5 \cdot \frac{1}{6} + (2) \cdot 10 \cdot \frac{1}{6} = \frac{30}{6} = \boxed{\$5 = E[R]}$$