Two basic types of random variables

Look at the set of all possible values of the RV.

Discrete random variable: A rv X is *discrete* if the set of all its possible values is countable.

Ex:

- X = # hits on the UD website in a day; possible values: $\{0, 1, \dots\}$ — infinite but countable set
- X = # days it rains in the month of July; possible values: $\{0, 1, 2, \dots, 31\}$
- X = binary outcome; possible values $= \{0, 1\}$. countable set

Continuous random variable: A rv X is *continuous* if the set of all possible values is uncountable, e.g., an interval.

Ex:

- X = height of a student (inches); possible values: $(0, x_{max})$
- X = highest temp. in the day in a city (°F); possible values: (x_{min}, x_{max})
- X = time b/w two successive landings at busy airport; possible values: (x_{min}, x_{max}).

Discrete X

Probability distribution: $f_\theta(x) = P[X=x]$ → $=0 \Rightarrow x$ is not a possible value; $>0 \Rightarrow x$ is a possible value

prob. mass function (pmf)

$\boxed{\sum_x f_\theta(x) = 1}$

Computing probability: $P(X \in A) = \sum_{x \in A} f_\theta(x)$

set

Interpretation of $P(E)$, probability of an event E

long-run proportion

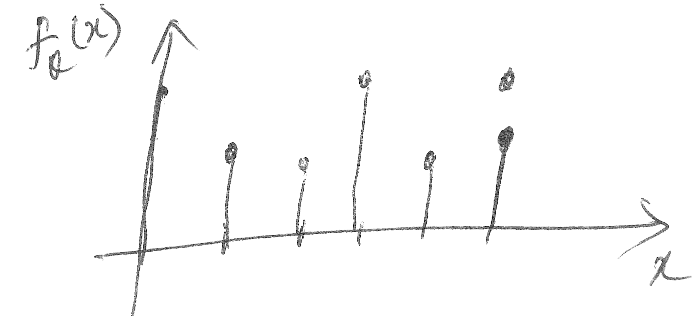
Think about repeating the experiment under identical conditions a large # times.

$$P(E) = \lim_{N\to\infty} \frac{\#\text{ times event } E \text{ occurs}}{\cancel{\text{total}}\ N}$$

N → # repetitions of the experiment

Expected value: $\boxed{E(X) = \sum_x \cancel{x}\, f_\theta(x)}$

"μ"



Interpretation of $E(X)$

- $E[X]$ = Average in the population that X represents
- $E[X]$ = center of the prob. dist. (or the balance point ~~of~~ of the prob. dist.)
- $E[X]$ = best guess at value of X
- $E[X] \cong$ long-run average value of X.

$$E[g(X)] = \sum_{x} g(x) f_X(x)$$

Discrete X (cont'd)

Variance: $\text{var}(X) = \overset{\sigma^2}{E(X-\mu)^2} = \sum_x (x-\mu)^2 f_\theta(x)$, $\mu = E(X)$

(verify): $\boxed{\text{var}(X) = E(X^2) - \mu^2}$

Standard deviation: $\underset{\uparrow\,\sigma}{\text{sd}(X)} = \sqrt{\text{var}(X)}$

Interpretation of $\text{var}(X)$ and $\text{sd}(X)$

- Measure how far the values of X are from 'μ'.
- ~~60~~ SD is easier to interpret than variance
- $\sigma^2 \geq 0$
- $\sigma = 0 \equiv P(X = \underset{\uparrow\,\text{constant}}{c}) = 1$, [In this case: $E(X) = c$].
- σ measures how far off the ~~[illegible]~~ best guess is.

Chebychev's inequality: $(k>0)$.

$$\underbrace{P\left[\,|X-\mu| > k\sigma\,\right]}_{\displaystyle \| \atop \displaystyle P\left[\,X > \mu - k\sigma \text{ or } X < \mu + k\sigma\,\right]} \leq \frac{1}{k^2}$$

$k=1$: $P\left[\,|X-\mu| > \sigma\,\right] \leq 1$. — not useful

$k=2$: $P\left[\,|X-\mu| > 2\sigma\,\right] \leq \frac{1}{4}$

$k=3$: $P\left[\,|X-\mu| > 3\sigma\,\right] \leq \frac{1}{9}$

$$\Downarrow$$

$$P\left[\,|X-\mu| \leq 3\sigma\,\right] \geq \frac{8}{9}$$

Often times: $[\mu - 3\sigma, \mu + 3\sigma]$ — "effective" range of X.

Continuous X

$P[X=x] = 0$ for all x.

Probability distribution: $f_\theta(x)$ $\cancel{=}$ $[\neq P[X=x]]$.

Probability density function (PDF)

$$\boxed{\int_{-\infty}^{\infty} f_\theta(x)\,dx = 1}$$

Computing probability: $\boxed{P(X \in A) = \int_A f_\theta(x)\,dx}$

Take any formula for the discrete case, replace 'Σ' by '$\int$', and 'pmf' by 'pdf', we get the corresponding formula for the cont. case.

Expected value: $E(X) = \int_{-\infty}^{\infty} x f_\theta(x)\,dx$

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) f_\theta(x)\,dx}$$

Variance: $\mathrm{var}(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_\theta(x)\,dx$

Interpretation:

same as in the discrete case.

$X \sim \text{Bernoulli}(p)$.

$$E[X] = \sum_x x f_p(x) = 1(p) + 0(1-p) = p.$$

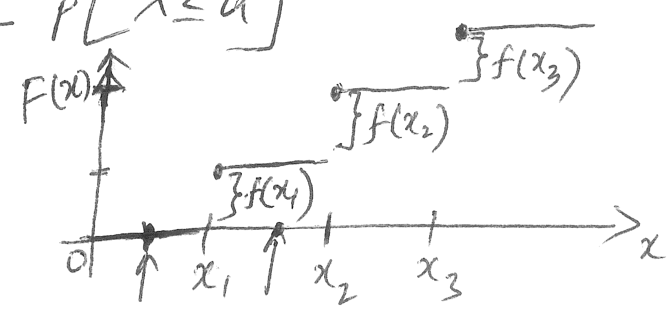
Cumulative distribution function of X

Cumulative distribution function (cdf) of X:

- $F(x) = P[X \le x]$, $x \in \mathbb{R}$
- Nondecreasing function of x
- One to one correspondence with pdf/pmf
- Plays a key role in simulation of random variables
- $\boxed{P(a < X \le b)} = F(b) - F(a)$
  $= P[X \le b] - P[X \le a]$

Discrete X with pmf $f(x)$

- $F(x) = P[X \le x]$, $x \in \mathbb{R}$
- Jump (or step) function of x



- *Getting pmf from cdf*: points of jump = possible values of X; sizes of jumps = probabilities

Continuous X with pdf $f(x)$

- $F(x) = P[X \le x] = \int_{-\infty}^{x} f_{\theta}(y)\,dy$
- Increasing function of x over the possible values of x.
- *Getting p~~m~~df from cdf:* $f_{\theta}(x) = \frac{d}{dx} F(x)$. (Fundamental thm. of calculus)

Some key model distributions

Discrete distributions

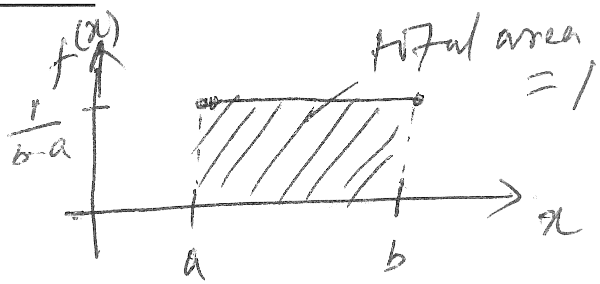
- $X \sim$ Bernoulli(p), Possible values: 0, 1 ; $f_p(x) = p^x(1-p)^{1-x}$, $x = 0, 1$
- $X \sim$ discrete uniform $(x_1, \ldots, x_N)$: $f_p(x_i) = \frac{1}{N}$, $i = 1, 2, \ldots, N$.
- $X \sim$ Binomial (n, p) ; $f_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$; $x = 0, 1, \ldots, n$

'Perform n indep. and identical Bernoulli trials resulting in outcomes $X_1, \ldots, X_n$. Then: $X = X_1 + \cdots + X_n$

$=$ # successes (1's) in the experiment.

Continuous distributions

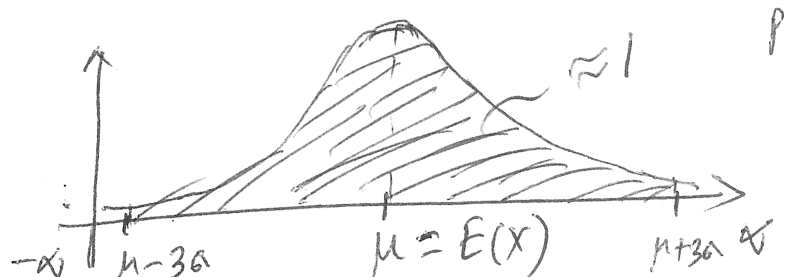
- $X \sim$ Uniform (a, b)



$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

- $X \sim N[\mu, \sigma^2]$

μ: $E(X)$, σ^2: $Var(X)$



$P[\mu - 3\sigma < X < \mu + 3\sigma] \approx 1$