Statistics Review - part 01

① Random Variables (RV's)

Definition: A random variable (RV) is a function from the sample space to the real line, i.e. $X: S \to \mathbb{R}$.

Notation: most commonly represented by capital letters (X, Y, Z, etc) whereas small letters (x, y, z, etc.) represent particular values (i.e. observations) of the RV.

Therefore, you can say: $P(X=x)$ as the probability of the RV X assume the value x.

Example: Flip two coins and suppose X is the RV corresponding to the # of heads.

- What is the sample space S? Just list all the possible outcomes

  $S = \{HH, HT, TH, TT\} = 2^2 = 4$ possible outcomes.

- What values can X assume? 0, 1 and 2.
- What are the probabilities associated with those values?

  $P(X=0) = 1/4$, $P(X=1) = 1/2$, $P(X=2) = 1/4$.

Example: in a ruler that goes from 0 to 1 meter, select a Real number at random. Suppose X is the RV corresponding to the measurement.

- What is the probability that $X = 0.1$ (i.e $P(X=0.1)$)?

  Answer: 0! Why? Infinite number of "equally likely" outcomes.

- What is $P(X \leq 0.5)$? Answer = 0.5
- What about $P(X \in [0.3, 0.7])$? Answer = 0.4.

→ What do the examples above tell us?

Differences about discrete RV's (first example) and continuous RV's (second example).

Definition: A discrete RV is one where the number of possible values is finite or countably finite. Otherwise, the RV is continuous (probability = 0 at every point).

1.1 Discrete RV's: if X is a discrete RV, its Probability Mass Function (pmf) is $f(x) = P(X = x)$

Important: $0 \leq f(x) \leq 1$ and $\sum_x f(x) = 1$

Recall our previous exemple & Flip 2 coins (X is # of heads)

Pmf is:

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0 \text{ or } x = 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

(Assumption: coin is fair, ie, prob(Heads) = prob(tails) $\Leftrightarrow p = 1/2$)

Or what if I want a more general pmf for n trials (# of coins) and p as the probability of a success (in this case, success is getting "head" as the outcome) with K successes ($n-k$ "failures")?

$$\underbrace{HHHH \cdots H}_{K \text{ successes (heads)}} \underbrace{TT \cdots T}_{(n-k) \text{ failures (tails)}} = p^K (1-p)^{n-K}$$

What is missing? How many ways can I arrange this sequence? $\binom{n}{K}$

So, a general pmf is $P(X = K) = \binom{n}{K} p^K (1-p)^{n-K} \sim \text{Binomial}(n, p)$

That pmf can be applied to many other situations, for example, for a fair dice, if I roll it 3 times, what is P(get exactly two 6's)?

This is a binomial dist. with parameters $n = 3$ and $p = 1/6$ (fair dice).

We want: $P(X=2)$ (X is a RV representing # of 6)

3 choose 2

$$P(X=2) = \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^{3-2} = \frac{15}{216}$$

Recall: $\binom{n}{K}$

$$\binom{n}{K} = \frac{n!}{K!(n-K)!}$$

More complex exemple: if I roll 2 dice 12 times. What is P(result will be 7 or 11 exactly 3 times)? X = # of times you get 7 or 11.

$P(7 \text{ or } 11) = P(7) + P(11) = \frac{6}{36} + \frac{2}{36} = 2/9 \Rightarrow X \sim \text{Bin}(12, 2/9)$

$$P(X=3) = \binom{12}{3} \left(\frac{2}{9}\right)^3 \left(\frac{7}{9}\right)^9 \quad \text{(Do the math!)}$$

(1.2) Continuous RV's: if X is a continuous RV, its Probability Density Function (pdf) satisfies:

i. $\int_{\mathbb{R}} f(x)dx = 1$ (area under $f(x)$ is 1)

ii. $f(x) \geq 0, \forall x$ (non-negative)

→ Also, if A is a subset of $\mathbb{R}$ (ie $A \subseteq \mathbb{R}$) $\Rightarrow P(X \in A) = \int_A f(x)dx$

→ You can think of $f(x)$ as $f(x)\,dx \approx P(x < X < x + dx)$

Remember: • $P(a < X < b) = \int_a^b f(x)dx$

• $P(X = x) = 0$

Notice the $\neq$ between (discrete) pmf and (continuous) pdf.

pmf: $f(x) = P(X = x)$
$0 \leq f(x) \leq 1$

vs

pdf $f(x)dx \approx P(x < X < x+dx)$
$f(x) \geq 0$ (possibly > 1)
↳ ex: $f(x) = 2(1-x)$ for $0 \leq x \leq 1$

→ One example: if X is "equally likely" to be anywhere between a and b, the $X \sim Uniform(a,b)$
↳ we say X is uniform distributed or X has uniform distribution

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

Let's check if $f(x)$ is really a pdf:

i. $\int_a^b f(x)dx = 1 \Rightarrow \int_a^b \frac{1}{b-a}dx = \frac{b-a}{b-a} = 1$ ✓

ii. $f(x) \geq 0, \forall x$ (notice that this holds even for a and b negative, since $a < x < b \Rightarrow b > a \Rightarrow \frac{1}{b-a} \geq 0$) ✓

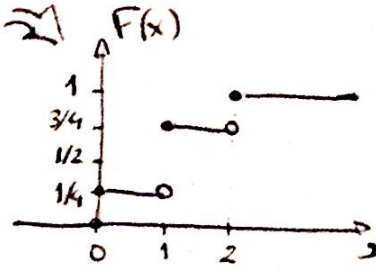
Since $\boxed{i}$ and $\boxed{ii}$ hold, $f(x)$ is a pdf.

② Cumulative distribution function (cdf): for any RV X (continuous or discrete) the cdf is defined as $F(x) = P(X \leq x)$

- if X is discrete: $F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} p(x_i)$
- if X is continuous: $F(x) = \int_{-\infty}^{x} f(t)\,dt$ (also $F'(x) = f(x)$ Proof by Fundamental Thm of Calculus)

→ Discrete cdf example: let's go back to our "Flip a coin 2x" example. X is the # of heads → $X = \begin{cases} 0 \text{ or } 2 & \text{with prob } 1/4 \\ 1 & \text{with prob } 1/2 \end{cases}$

The cdf $F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \leq x < 1 \\ 3/4 & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$ ⇒



Remember $X \sim Bin(n=2, p=1/2)$, so the following formula applies

$$P(X \leq 1) = \sum_{x_i \leq 1} P(X = x_i) = \sum_{i=0}^{1} \binom{2}{i} \left(\frac{1}{2}\right)^i \left(1-\frac{1}{2}\right)^{2-i} = \underbrace{\binom{2}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^2}_{} + \underbrace{\binom{2}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^1}_{}$$

(Be careful with ≤ or < with discrete RV)

$\therefore\ P(X \leq 1) = \frac{1}{4} + \frac{1}{2} = 3/4$ (the exact same answer)!

→ Continuous cdf example: $X \sim U(0,1)$

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \Rightarrow F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

for a specific interval inside $(0,1)$, say $(0.2, 0.5)$

$$\int_{0.2}^{0.5} 1\,dx = x\Big|_{0.2}^{0.5} = 0.3$$ or equivalently

$$F(0.5) - F(0.2) = \int_0^{0.5} 1\,dx - \int_0^{0.2} 1\,dx = x\Big|_0^{0.5} - x\Big|_0^{0.2} = 0.5 - 0.2 = 0.3$$ //

③ Expected value of a RV: the mean or expected value or average of a RV X is

$$\mu = E[X] = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Example: Roll a fair dice. Let X be a RV representing the possible outcome (i.e. $X = 1, 2, \cdots, 6$ each with prob. $1/6$).

$$E[X] = \sum_x x f(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Example: Flip a coin, with p being the prob. of getting a head and $q\ (= 1-p)$ the prob. of getting a tail. Let X be the RV representing the # of heads.

$$E[X] = \sum_x x f(x) = 0 \cdot q + 1 \cdot p = p \quad \therefore X \sim \text{Bernoulli}(p)$$

Example: $X \sim U(0,1)$.

$$E[X] = \int_{\mathbb{R}} x f(x) dx = \int_0^1 x(1) dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2} - \frac{0}{2} = \frac{1}{2}$$

LOTUS: Law of the Unconcious Statistician

Definition/Theorem: The expected value of a function of X, say h(X) is

$$E[h(X)] = \begin{cases} \sum_x h(x) f(x) = \sum_x h(x) P(X=x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} h(x) f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Think of $E[h(X)]$ simply as a weighted function of $h(x)$, with the weights being $f(x)$ values.

Examples: $E[X^2] = \int_{\mathbb{R}} \overbrace{x^2}^{h(x)} f(x) dx \qquad E[\sin X] = \int_{\mathbb{R}} (\sin x) f(x) dx$

k^{TH} central moment $\Leftarrow$ $E[(X-\mu)^k] = \begin{cases} \sum_x (x-\mu)^k P(X=x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} (x-\mu)^k f(x) dx & \text{if } X \text{ is continuous} \end{cases}$

$E[aX + b] = aE[X] + b \qquad E[g(X) + h(X)] = E[g(X)] + E[h(X)]$

4 Variance, Covariance and Correlation

4.1 • Variance is the second central moment, i.e. $var(X) = E[(X-\mu)^2]$
It's a measure of spread or dispersion.

→ Theorem. $var(X) = E(X^2) - (E[X])^2$

<u>Proof</u>: $Var(X) = E[(X-\mu)^2] = E[X^2 - 2X\mu + \mu^2] = E[X^2] - 2\mu \overbrace{E[X]}^{\mu} + \mu^2$

$\Leftrightarrow Var(X) = E[X^2] - \mu^2 = E[X^2] - (E[X])^2$ //

<u>Example</u>: $X \sim U(a,b) \Rightarrow f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

$E[X] = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$ $\quad E[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{a^2+ab+b^2}{3}$

$Var(X) = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(a-b)^2}{12}$

→ Theorem. $Var(aX+b) = a^2 Var(X)$

<u>Proof</u>: $var(aX+b) = E[(aX+b)^2] - [E(aX+b)]^2$

$= E[a^2X^2 + 2abX + b^2] - (aE[X]+b)^2$

$= a^2E[X^2] + \cancel{2abE[X]} + b^2 - (a^2(E[X])^2 + \cancel{2abE[X]} + \cancel{b^2})$

$= a^2E[X^2] - a^2(E[X])^2 = a^2(E[X^2] - (E[X])^2) = a^2 var(X)$ //

<u>Example</u> $Y = 4X + 5$.

$var(Y) = 16\, var(X)$.

4.2. covariance: the covariance between two RV's X and Y is

$cov(X,Y) \equiv \sigma_{XY} \equiv E[(X-E[X])(Y-E[Y])]$ (implies $cov(X,X) = E[(X-E[X])^2] = var(X)$)

$= E[XY - XE[Y] - YE[X] + E[X]E[Y]]$

$= E[XY] - E[X]E[Y] - \cancel{E[Y]E[X]} + \cancel{E[X]E[Y]}$

$cov(X,Y) = E[XY] - E[X]E[Y]$

$cov(X,Y) = \begin{cases} \oplus & X, Y \text{ move "in the same direction" (ex: height \& weight)} \\ \ominus & X, Y \text{ move "in opposite directions" (ex: snowfall \& temperature)} \end{cases}$

if X and Y are indep then

$cov(X,Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$

X, Y indep implies $cov(X,Y) = 0$, but the opposite is NOT true!

Example: $X \sim U(-1,1)$ and $Y = X^2$ (X, Y dependent)

$E[X] = \int_{-1}^{1} x \cdot \frac{1}{2}\, dx = 0$

$E[XY] = E[X^3] = \int_{-1}^{1} x^3 \frac{1}{2}\, dx = 0$

$\text{cov}(X,Y) = E[XY] - E[X]E[Y] = 0$

$\text{cov} = 0 \nRightarrow$ dependent RV's

Also, correlation does not imply causation.
(check Buzzfeed: The most 10 bizarre correlations).

4.3 correlation between X and Y is

$$\rho = \text{Correl}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Obs: cov has "square" units, corr is unitless.

Corollary: X, Y indep implies $\rho = 0$

Theorem: $-1 \le \rho \le 1$ (then bounds for $\text{cov}(X,Y)$ can be calculated $-1 \le \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \le 1$

$\rho \approx 1$ (high $\oplus$ corr)

$\rho \approx 0$ (low corr)

$\rho \approx -1$ (high $\ominus$ corr)

$\therefore -\sigma_X\sigma_Y \le \text{cov}(X,Y) \le \sigma_X\sigma_Y$ → product of std. deviation

4.4 Some theorems

→ $\text{Var}(X+Y) = \text{var}(X) + \text{var}(Y) + \underbrace{2\,\text{cov}(X,Y)}_{=0 \text{ if } X,Y \text{ are indep.}}$ (whether X, Y indep or not)

proof: $\text{var}(X+Y) = \underbrace{E[X^2] - (E[X])^2}_{=\text{var}(X)} + \underbrace{E[Y^2] - (E[Y])^2}_{\text{var}(Y)} + \underbrace{2(E[XY] - E[X]E[Y])}_{2\,\text{cov}(X,Y)}$

$= \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X,Y)$

→ $\text{var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{var}(X_i) + 2\sum\sum_{i<j} \text{cov}(X_i, X_j)$

→ $\text{cov}(aX, bY) = ab\,\text{cov}(X,Y)$

→ $\text{var}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2\,\text{var}(X_i) + 2\sum\sum_{i<j} a_i a_j\,\text{cov}(X_i, X_j)$

Example: $\text{var}(X-Y) = \text{var}(X) + \text{var}(Y) - 2\,\text{cov}(X,Y)$

Example: $\text{var}(X-2Y+3Z) = \text{var}(X) + 4\,\text{var}(Y) + 9\,\text{var}(Z) - \overbrace{4}^{2.2.1}\text{cov}(X,Y) + \overbrace{6}^{2.1.3}\text{cov}(X,Z) - \overbrace{12}^{2.2.3}\text{cov}(Y,Z)$