Lecture Slides for

EXPECTATION AND LAW OF LARGE NUMBERS

Dr. LALIT KUMAR SINGH

Expectation

- Random variables, by definition, result in different outcomes
- The variation in random variables is captured by their **distribution**
  - Probability Mass function for Discrete variables
  - Probability Density function for Continuous variables
- We use summary statistics such as ***expectation*** (mean) and ***variance*** to capture some overall properties of the distribution/variable
- **Expectation** gives mean/average/expected value of the random variable given the distribution
  - E.g : Expected returns on a certain investment in the market
  - E.g : Expected rainfall during coming monsoon

Expectation

- The **expectation**, or **expected value**, of some function $f(x)$ with respect to a probability distribution $P(x)$ is the average value of $f(x)$ when x is drawn from P $[x \sim P]$

Denoted by $\mathbb{E}_{x\sim P}[f(x)]$

- If P is clear from the context $\mathbb{E}_x[f(x)]$
- If x is also clear from the context $\mathbb{E}[f(x)]$
- Sometimes, simply denoted as $\mathbb{E}[f]$

- Mathematically,

Discrete $$\mathbb{E}_{x\sim P}[f(x)] = \sum_x \overset{\text{p.m.f}}{P(x)} f(x)$$

Continuous $$\mathbb{E}_{x\sim P}[f(x)] = \int_x \underset{\text{p.d.f}}{p(x)} f(x)dx$$

Multivariate Expectation

vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

- For a multivariate random variable x we can interpret the variable and the expectations by considering each component separately

That is, if $x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_D \end{bmatrix} \in \mathbb{R}^D$ then

$\vec{x} = \begin{bmatrix} \text{Temp} \\ \text{Press} \\ \text{Humidity} \end{bmatrix} \quad \begin{bmatrix} E_{temp}[f(temp)] \\ E_{press}[f(press)] \\ \vdots \end{bmatrix}$

$$\mathbb{E}_x[f(x)] = \begin{bmatrix} \mathbb{E}_{x_1}[f(x_1)] \\ \mathbb{E}_{x_2}[f(x_2)] \\ \dots \\ \mathbb{E}_{x_D}[f(x_D)] \end{bmatrix}$$

Examples

- What is the expected value of a coin toss for a fair coin assuming that Heads=1 and Tails=0?

x	0	1
$P(x)$	$1/2$	$1/2$

Ans: Random variable $X \in \{0,1\}$.

Identify random var

Calculate p. distribution

P is a ***uniform*** distribution with both states having probability ½

So, $\mathbb{E}_{x\sim P}[x] = \sum_x xP(x) = \left[0 \times \frac{1}{2} + 1 \times \frac{1}{2}\right] = \frac{1}{2}$

- Similarly, the expected value of a fair dice throw is?

Random variable $X \in \{1,2,3,4,5,6\}$. P is uniform with probability $\frac{1}{6}$

So, $\mathbb{E}_{x\sim P}[x] = \sum_x xP(x) = \left[1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots 6 \times \frac{1}{6}\right] = 3.5$

Examples

- What is the expected value of the sum of two dice thrown together?

Ans: Random variable $x \in \{2,3, \dots, 12\}$.

What is the probability distribution?

- Note : P is not uniform

Distribution

x	2, 12	3, 11	4, 10	5, 9	6, 8	7
P(x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$

So, $\mathbb{E}_{x \sim P}[x] = \sum_x xP(x) = \left[2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \cdots 12 \times \frac{1}{36}\right] = 7$

Question : Is there an easier way of calculating this case?

Linearity of Expectation

$E[\cdot]$

- Important Property of expectation
  The Expectation Operator is linear

  Linear Comb

- Mathematically, if $f(x) = \alpha g(x) + \beta h(x)$ is a multivariate function with $\alpha, \beta \in \mathbb{R}$ being scalars, then

$$\boxed{\mathbb{E}[f] = \alpha\mathbb{E}[g] + \beta\mathbb{E}[h]}$$

Note the use of compact notation

Applying this to our example, we note that $X = D_1 + D_2$ where D_1 and D_2 are the number obtained on the first and second dice respectively.

$E[X] = E[D_1+D_2] = E[D_1]+E[D_2]$

Then, $\mathbb{E}[X] = \mathbb{E}[D_1] + \mathbb{E}[D_2] = \underline{3.5} + 3.5 = 7$

Note : Much simpler, since the distribution of X need not be found

Proof of Linearity of Expectation

Linearity: If $f(x) = \underbrace{\alpha g(x) + \beta h(x)}$ is a multivariate function with $\alpha, \beta \in \mathbb{R}$ being scalars, then

$$\mathbb{E}[f] = \alpha\mathbb{E}[g] + \beta\mathbb{E}[h] \rightarrow \text{Claim}$$

Proof: For continuous distributions

$$
\begin{aligned}
\mathbb{E}[f] &= \int f(x)p(x)dx \quad \rightarrow \text{Definition} \qquad \mathbb{E}[f] = \sum f(x)p(x) \\
&= \int \big(\alpha g(x) + \beta h(x)\big)p(x)dx \\
&= \alpha \overbrace{\int g(x)p(x)dx}^{\mathbb{E}[g]} + \beta \int \underbrace{h(x)p(x)dx}_{\mathbb{E}[h]} \\
\Rightarrow \mathbb{E}[f] &= \alpha\mathbb{E}[g] + \beta\mathbb{E}[h]
\end{aligned}
$$

Discrete can be proved similarly – Try it as an exercise!

Law of large numbers

Law of large No.

$E(x) = 0.5$

	H	T	
$10 \rightarrow$	2	8	0.2
$100 \rightarrow$	40	60	0.4
$1000 \rightarrow$	550	450	0.55

Mixture distribution

- Mixture distribution is a way of combining many distributions.
- $c_1 = \mathcal{N}(x; \mu_1, \sigma_1^2), c_2 = \mathcal{N}(x; \mu_2, \sigma_2^2)$.
- Prior probability: $\alpha_1 = P(c = c_1), \alpha_2 = P(c = c_2)$
- Posterior Probability: $P(c|x)$
- Mixture distribution: $P(x) = \sum_i P(c = c_i) P(x|c)$.