Random Variables

└→ Discrete – PMF, CMF (Σ)
└→ Continuous – PDF, CDF ($\int$)

- A Random Variable is a Numerical quantity that results from a random phenomenon.

Discrete R.V

PMF

$$P(X=x) = P_X(x)$$

CMF

Continuous R.V

PDF

$$F(a < X < b) = \int_a^b f_X(x)\,dx$$

CDF

In machine learning, R.V is often used
- to model uncertainty in predictions
- Hypothesis testing
- Regression analysis

└→ Statistical process for estimating the relationships between dependant and independent variables

Multivariate Probability Distributions

- Describing the probability Outcomes for mutiple random variables

$$P(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) \qquad \begin{matrix} -\infty < y_1 < \infty \\ -\infty < y_2 < \infty \end{matrix}$$

Joint Probability description

- Describes the probability of each combination of outcomes of multiple random variable.

Marginal Probability Distribution

- The probability distribution of a subset of the random variable.

$$P(X_1) = \int P(X_1, X_2)\, dx_2$$

Common Multivariate distributions

→ Multivariate normal distribution
(Generelization of normal distribution)

→ Multinomial distribution
(Generelization of binomial distribution)

$$F(u,v) = P(U \le u, V \le v)$$

$$= \int_{-\infty}^{u}\int_{-\infty}^{v} f(t_1, t_2)\, dt_2 \, dt_1$$

Common features multivariate distributions.

→ Correlation

→ Independence $P(x_1, x_2) = P(x_1)\,P(x_2)$

Example Usages of Mutivariate distributions

→ ANOVA (Analysis of varience)

- Statistical method used to test differences between two / more means.
- IF the between-group variance is high and within-group variance is low, this provides evidence that the means of the groups are significantly different.

Conditional Probability

– Probability of an event happening given that another cvent has already happened.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Joint PDF

$$f(u|v) = \frac{f_{uv}(u,v)}{f_v(v)}$$

$$f_v(v) = \int_{-\infty}^{\infty} f_{uv}(u,v)\, fu$$

Marginal PDF

Important Rules of Probability

$$P(X=x_i) = \sum_j P(X=x_i, Y=y_j) \leftarrow \text{Sum rule}$$

$$P(X=x_i, Y=y_i) = P(Y=y_i \mid X=x_i)\, P(X=x_i)$$

Product rule

$$P(X=x_i) = \sum_j (X=x_i \mid Y=y_j) \cdot P(Y=y_j)$$

The law of total Probability

Bayes Theorem

→ Allows to invert the conditional probability

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Frequentist View

- Probabilities are defined as the long-run relative frequency of an event occurring after many repetitions of an experiment.

- Model parameters are Fixed but Unknown values.

- Inferencing will estimate these fixed params.

- Focuses on estimating parameters or testing hypothesis without incorporating prior beliefs.

- Confidence interval and P-Values are central Concept.

- There will be only one-true value for a given parameter; but this value will change with the sampling distribution.

- These lead us to uncertainity in parameters

Subiective Problistic Intepretation

/ /

Bayesian View

- Intepreted as a degree of belief about an event based on current knowledge or evidence.

- Model parameters are considered random variables with probability distributions that reflect our uncertainity about the true values

- Inference involves updating a prior-belief (prior-distribution) using observed data to form a posterior distribution.

- Uncertainity about the parameter is directly modeled through the posterior distribution, Combining prior belief and evidence from data.

Subjective probability

→ Contains no formal calculations
→ Derived from an individuals personal judgement about whether a specific outcome is likely to occur.

Bayesian Inference

→ A method in statistics and Machine learning that applies Bayes' theorem to update the probability estimate for a hypothesis as new evidence or data available. (Rather than providing a single estimate, giving a probability distribution over all possible values)

$$\underbrace{P(\theta/x)}_{\text{posterior}} = \frac{\overbrace{P(x|\theta)}^{\text{Likelihood}} \cdot \overbrace{P(\theta)}^{\text{prior}}}{\underbrace{P(x)}_{\text{evidence}}}$$

Prior [$P(\theta)$]

→ Belief about the parameter θ before we see the data.

describes how likely it is to observe the data for given params.

Likelihood [$P(x|\theta)$]

→ How well the parameter θ explains the observed data.

Posterior [$P(\theta|x)$]

→ Updated belief about the parameter θ after taking into account the observed X

Posterior $\propto$ Likelihood $\times$ Prior.

Evidence $P(x)$

→ A normalizing factor that ensures the posterior is a proper distribution function

→ Marginal likelihood function

$$p(x) = \int p(x|\theta)\, p(\theta)\, d\theta$$

More insights on likelihood.

→ How likely it is to observe the given data assuming a particular model and it's parameters are correct.

→ "Given that this event has occurred, how likely is it that a particular parameter value caused it"

$$L(\theta|D) = P(D|\theta)$$

→ Likelihood will be a function of parameter.

→ evaluates how well different parameters explain observed data

Minimum Likelihood Estimator.

→ Parameter values that maximize the likelihood of the observed data.

$$\hat{\theta}_{MLE} = \text{argmax}_{\theta} \; P(D|\theta)$$

Maximum Aposteriori Estimator.

→ Seeks to find the parameter value that best explains the observed data by incorporating prior beliefs or knowledge about the parameters.

$$\hat{\theta}_{MAP} = \text{argmax}_{\theta} \; P(\theta|D)$$

$$= \text{argmax}_{\theta} \; P(D|\theta) \, P(\theta)$$

→ As the data size increases, MAP converges to MLE (IF prior is weak)

Check?

4. Example: Coin Toss Problem

Let's say we want to estimate the probability of heads p in a biased coin. We observe 8 heads out of 10 tosses.

Step 1: Choose a Prior

Assume a **Beta prior distribution** for p:

$$P(p) = \text{Beta}(2, 2)$$

This represents a weak belief that the coin is fair (since the Beta(2,2) is symmetric and centered at 0.5).

Step 2: Likelihood Function

Given the observed data (8 heads in 10 tosses), the likelihood follows a binomial distribution:

$$P(X \mid p) = \binom{10}{8} p^8 (1-p)^2$$

Step 3: Posterior Distribution

The posterior distribution combines the prior and likelihood:

$$P(p \mid X) \propto P(X \mid p) \cdot P(p)$$

This gives a new **Beta distribution** as the posterior:

$$P(p \mid X) = \text{Beta}(10, 4)$$

This posterior reflects our updated belief about p, suggesting a higher probability of heads based on the data.

Step 4: Inference from the Posterior

From the posterior distribution $\text{Beta}(10, 4)$, we can:

- Calculate the **mean** of p (expected value) as:

  $$\mathbb{E}[p] = \frac{10}{10+4} = 0.71$$

  So, the expected probability of heads is around 0.71.

- Find a **credible interval** for p, say the 95% interval. This tells us that with 95% confidence, p lies within a certain range.