Important Concepts in CS 215

Anish Kulkarni

IIT Bombay

September 2024

Sample Points

VS

Probability Distribution

Sample Points

Let $x_1, x_2, \cdots, x_N$ be a sample set.

$$\text{Mean } \mu = \bar{x} = \frac{1}{N}\sum_{i=1}^{n} x_i$$

$$\text{Variance } \sigma^2 = s^2 = \frac{1}{N-1}\sum_{i=1}^{n}(x_i - \bar{x})$$

Chebyshev's Inequality

$$\frac{\#\{i : |x_i - \bar{x}| > ks\}}{N} < \frac{1}{k^2}$$

If $y_1, y_2, \cdots, y_n$ is another sample set then

$$\text{Correlation Coeffecient } r = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{(N-1)\sigma_x\sigma_y}$$

Probability Distribution

We consider a sample space Ω. An **Event** A is any[1] subset of Ω.
Let $\mathcal{P}(\Omega)$ denote the set of all subsets of Ω.
A probability function $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ satisfies the following axioms:

1. $\mathbb{P}(\Omega) = 1$
2. If $A_1, A_2, \cdots$, form a countable collection of disjoint subsets of Ω then

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

[1] If the sample space is not finite, but a continuous range, we need to restrict event to special kinds of subsets called measurable sets

Random Variable and CDF

A random variable X is simply a function from Ω to $\mathbb{R}$.
Using this we can extend the definition of $\mathbb{P}$ to subsets U of $\mathbb{R}$ as follows

$$\mathbb{P}(U) = \mathbb{P}(X^{-1}(U))$$

In particular if U is a singleton $\{x\}$ we denote $\mathbb{P}(U)$ by $\mathbb{P}(X = x)$

Let X be a random variable, we define its **CDF**[2] $F : \mathbb{R} \to \mathbb{R}$ by

$$F(c) = \mathbb{P}([-\infty, c]) = \mathbb{P}(X \leq c)$$

In case the CDF turns out to be a differentiable function, we define the **PDF**[3] by

$$f_X(x) = F'(x)$$

[2] Cumulative Distribution Function

[3] Probability Density Function

Expectation of a Random Variable

Let X be a random variables and $g : \mathbb{R} \to \mathbb{R}$ be any function

- Case 1: X is discrete, that is takes finitely many values $u_1, u_2, \cdots, u_n$.

$$\mathbb{E}(g(X)) = \sum_{u=0}^{n} g(u_i)P(X = u_i)$$

- Case 2: X is continuous (for now assume this means it has a pdf). Then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

Linearity of Expectation

$$\mathbb{E}[\sum_{i=1}^{n}(a_i g_i(X_i) + b_i)] = \sum_{i=1}^{n}(a_i \mathbb{E}[g_i(X_i)] + b_i)$$

Variance

$$\mathrm{Var}(X) = \sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Random Variable Examples

Name	Formula	Mean	Variance
Bernoulli	$xp + (1-x)(1-p)$	p	$p(1-p)$
Binomial	$\binom{n}{k} p^k (1-p)^n$	np	$np(1-p)$
Geometric	$(1-p)^{n-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Poisson	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ
Uniform(α, β)	$\frac{1}{\beta-\alpha}$	$\frac{\alpha+\beta}{2}$	$\frac{(\alpha-\beta)^2}{12}$
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Exponential	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Beta	$\frac{1}{B(a,b)} x^{a-1} y^{b-1}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$

Table: Random Variables

Entropy and MGF

- **Continuous**

$$Ent(X) = -\int_{-\infty}^{+\infty} f_X(x)\log(f_X(x))dx$$

- **Discrete**

$$Ent(X) = -\sum_{i=1}^{n} P(X = x_i)\log(P(X = x_i))$$

Moment Generating Function:

$$\text{MGF}(X) = \phi(t) = \mathbb{E}[e^{Xt}]$$

Estimators

Given a sample set $D = x_1, x_2, \cdots, x_n$ sampled from a distribution X, which is a known distribution, we want to estimate the parameter.

Maximum Likelihood Estimator

$$MLE(D) = \hat{\theta} = \arg\max_{\theta} \mathbb{P}(D|\theta) = \arg\max_{\theta} \prod_{i=1}^{n} \mathbb{P}(x_i|\theta)$$

The estimator is called unbiased if $\mathbb{E}(\hat{\theta}) = \theta$. Where θ is the true parameter from which D was sampled.
It is called consistent if $\hat{\theta} \to \theta$ as $N \to \infty$

If we have a prior on θ then we can use Bayesian Estimator:

$$\hat{\theta} = \arg\max_{\theta} \mathbb{P}(\theta|D) = \arg\max_{\theta} \mathbb{P}(\theta) \times \mathbb{P}(D|\theta)$$