In the name of God, the Compassionate, the Merciful

STATISTICS

Statistical Review

-
- Continuous RV: Probability Density Function (PDF)
- Both: Cumulative Distribution Function (CDF)

Independence:

- $P(A \cap B) = P(A)P(B)$
- $P(A|B) = P(A)$
- Mutual Independence:
  - $P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2)\ldots P(A_k)$
  - $P(A_1, A_2, \ldots, A_k|B) = \prod_{i=1}^{k} P(A_i|B)$

Bayes' Theorem $P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$

Expected Value: Expected value is just a weighted average of the values of X, weighted by their probabilities.

- $\mathrm{E}[b] = b$ $\qquad$ $\mathrm{E}[aX + b] = a\,\mathrm{E}[X] + b$
- $\mathrm{E}[XY] = \mathrm{E}[X] * \mathrm{E}[Y]$ If X and Y are independent
- $\mathrm{E}[g(X)] = \int g(X)f(x)$

Variance & Standard Deviation:

- $\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}[(X-\mu)^2]$ $\qquad$ $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - \mu^2$
- $\mathrm{Var}(b) = 0$ $\qquad$ $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$
- $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ If X and Y are independent

Covariance:

- $\mathrm{Cov}(a + bX, c + dY) = bd\,\mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X, Y) = 0$ If X and Y are independent
- $\mathrm{Cov}(X, Y) > 0$ If X and Y have direct relationship
- $\mathrm{Cov}(X, Y) < 0$ If X and Y have inverse relationship
- $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y)$
- $\mathrm{Var}(X + Y + Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) + 2\,\mathrm{Cov}(X, Y) + 2\,\mathrm{Cov}(X, Z) + 2\,\mathrm{Cov}(Z, Y)$

Covariance Matrix (Σ):

$$\Sigma = \mathrm{E}[(X - \mathrm{E}[X])(X - \mathrm{E}[X])^T]$$

$$= \begin{bmatrix} \mathrm{E}[(x_1-\mu_1)^2] & \ldots & \mathrm{E}[(x_1-\mu_1)(x_n-\mu_n)] \\ \vdots & \ddots & \vdots \\ \mathrm{E}[(x_n-\mu_n)(x_1-\mu_1)] & \ldots & \mathrm{E}[(x_n-\mu_n)^2] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \ldots & \sigma_{nn} \end{bmatrix} \text{ where } \quad \sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$$

- Σ is positive-semidefinite and symmetric
- $\Sigma = \mathrm{E}[XX^T] - \mu\mu^T$
- $\mathrm{Var}(AX + a) = A\,\mathrm{Var}(X)A^T$ $\qquad$ $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)^T$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(AX + a, BY + b) = A\,\mathrm{Cov}(X, Y)B^T$
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

Correlation Coefficient:

- $\mathrm{Corr}(X, Y) = \rho_{X,Y} = \dfrac{\mathrm{Cov}(X, Y)}{\sigma_X\sigma_Y}$, $-1 \le \rho \le 1$
- $\mathrm{Corr}(aX + b, cY + d) = \mathrm{Corr}(X, Y)$
- $\rho = \mp 1$ indicating perfect negative/positive association.

Correlation Matrix (R):

$$\Sigma = \Gamma\,\mathrm{R}\,\Gamma,\ \Gamma = \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n \end{bmatrix} \text{ and } \mathrm{R} = \begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{12} & 1 & \ldots & \\ \vdots & & \ddots & \vdots \\ \rho_{1n} & & \ldots & 1 \end{bmatrix}$$

Moments

- n^{th} order moment of a RV X is the expected value of X^n: $M_n = \mathrm{E}[X^n]$
- Normalized form (Central Moment): $M_n = \mathrm{E}[(X - \mu_X)^n]$.
- 1st moment is **Mean**.
- 2nd moment is **Variance**.
- 3rd moment is **Skew** (Measure of asymmetry)
- 4th moment is **Kurtosis** (Measure of flatness)

Marginal distribution: distribution of one or more variables $P(A) = \sum_B P(A, B)$

consistent estimator is an estimator having the property that as the number of data points used increases indefinitely, the resulting sequence of estimates converges in probability to θ

$$(??) \lim_{n\to\infty} \hat{\theta}_n = \theta \qquad \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

Point Estimation

- A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data.
- A measure of accuracy is the expected or mean square error (MSE): $E[(\hat{\theta} - \theta)^2]$
- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if $E(\hat{\theta}) = \theta$.
- The difference $E(\hat{\theta} - \theta)$ is called the bias of $\hat{\theta}$.
- maximum likelihood method, moments method

Interval estimation

- confidence intervals

Stochastic

Probability Space

1. Ω: The set of all possible outcomes which is called **Sample Space**.
2. ω: Every element Ω of is called an **Event**.
3. P: A **probability measure** intends to be a function defined for all subsets of Ω.

- Two events are mutually exclusive or **disjoint** if they cannot occur at the same time.
- Two events are **independent** if the occurrence of one does not change the probability of the other occurring.

Random Variables (RV): A random variable is a quantity whose value is determined by the outcome of an experiment.

Discrete Sample Spaces

Discrete RV: Probability Mass Function (PMF)

$$p_X[x] = P[X = x]$$

Bernoulli

$$p_X(k) = \mathrm{p};\ E[X] = \mathrm{p};\ Var[X] = \mathrm{pq}$$

- Any random variable whose only possible values are 0 and 1 occurs with probability p is called a **Bernoulli** random variable.
- The Bernoulli process consists of repeated independent Bernoulli trials with the same parameter p.

Binomial distribution

$$p_X(k) = \binom{n}{k}\mathrm{p}^k\mathrm{q}^{n-k};\ E[X] = n\mathrm{p};\ Var[X] = n\mathrm{pq}$$

- If you ask how many successes there will be among n Bernoulli trials, then the answer will have a **Binomial distribution**.
- **Examples**:
  - The number of heads/tails in a sequence of coin flips
  - Vote counts for two different candidates in an election
  - The number of male/female employees in a company
  - The number of accounts that are in compliance or not in compliance with an accounting procedure
  - The number of successful sales calls
  - The number of defective products in a production run
  - The number of days in a month your companys computer network experiences a problem
- The binomial probability law can be generalized to the case where we note the occurrence of more than one event, then the answer will have a **Multinomial distribution**. ... The vector specifies the $(k_1, \ldots, k_M)$ number of times each of the events B_i occurs.
- If you ask how many trials it will be to get the first success, then the answer will have a **Geometric distribution**.
- If you ask how many trials there will be to get the r^{th} success, then the answer will have a **Negative Binomial distribution**.
- Given that there are M successes among N trials, if you ask how many of the first n trials are successes, then the answer will have a **Hypergeometric**.

Continuous Sample Spaces

The **characteristic function** of a random variable X is defined by $\Phi_X(\omega) = E[e^{j\omega X}]$

Poisson distribution

- A **Poisson** process is the continuous version of a Bernoulli process.
- Whereas in a Bernoulli process either no or one event occurs in a unit time interval, in a Poisson process any nonnegative whole number of events can occur in unit time.
- If you ask how many events occur in an interval of length t, then the answer will have a **Poisson distribution**.
- It usually is applicable in situations where random "events" occur at a certain *rate* over a period of *time*.
- Examples:
  - The hourly number of customers arriving at a bank
  - The daily number of accidents on a particular stretch of highway
  - The hourly number of accesses to a particular web server
  - The daily number of emergency calls in Dallas
  - The number of typos in a book
  - The monthly number of employees who had an asence in a large company
  - Monthly demands for a particular product

Exponential distribution

- If you ask how long until the first event occurs, then the answer will have an **Exponential distribution**.
- typically used to model time intervals between "random" events
- Examples:
  - The length of time between telephone calls
  - The length of time between arrivals at a service station
  - The life time of electronic components, i.e., an inter-failure time

Gamma distribution

- If you ask how long until the r^{th} event, then the answer will have a **Gamma distribution**.
- For example, it is used to model the time required to service customers in queueing systems, the lifetime of devices and systems in reliability studies, and the defect clustering behavior in VLSI chips.
- If there are $\alpha + \beta$ events in a given time interval, if you ask what fraction of the interval it takes until the α^{th} event occurs, then the answer will have a **Beta distribution**.

Distributions related to the central limit theorem

- The Central Limit Theorem says sample means and sample sums approach normal distributions as the sample size approaches infinity.
- Normal distribution
- χ^2-distribution: ChiSquared(ν), The parameter ν, the number of "degrees of freedom"
- Students T-distribution
- Snedecor-Fishers F-distribution

References:
[1] ???
Made by ma.mehralian using LaTeX