Descriptive Statistics

Sample Mean and Sample Variance

Definition. *Sample Mean*

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

$$y_i = ax_i + b \Rightarrow \bar{y} = a\bar{x} + b$$

Given a frequency table with values v_i and frequencies f_i:

$$\bar{x} = \sum_{i=1}^{k} v_i f_i/n$$

Definition. *Sample Variance*

$$s^2 = \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)/(n-1)$$

$$y_i = ax_i + b \Rightarrow s_y^2 = as_x^2$$

Sample Percentiles

With data x in increasing order:

$$100p \text{ Percentile} = \begin{cases} \lceil x_{np} \rceil & , np \text{ is not an integer} \\ \frac{1}{2} x_{np} x_{np+1} & , \text{otherwise} \end{cases}$$

Definition. *Sample Correlation Coefficient*

$$\frac{1}{(n-1)s_x s_y} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Elements of Probability

Sample Space and Events

Events

Definition. *The sample space, S is the set of all possible outcomes. Subsets of S are individual events.*

Algebra of Events

Communative law:

$$E \cup F = F \cup E, EF = FE$$

Associative law:

$$(E \cup F) \cup G = E \cup (F \cup G), (EF)G = E(FG)$$

Distributive law:

$$(E \cup F)G = EG \cup FG, EF \cup G = (E \cup G)(F \cup G)$$

Axioms of Probability

Proposition.

$$1 = P(S) = P(E \cup E^{\complement}) = P(E) + P(E^{\complement})$$

$$P(E^{\complement}) = 1 - P(E)$$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Sample Spaces With Equally Likely Outcomes

$$P(E) = \frac{\text{Number of points in } E}{N}$$

n items picked k at a time without replacement when order matters, there are $\binom{n}{k}$ possible groups. For orderings, there are $n!$ ways to order all of the items.

Conditional Probability

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Bayes' Formula

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^{\complement})P(A^{\complement})}$$

Independent events

Definition. *E and F are independent events if:*

$$P(EF) = P(E)P(F)$$

Definition. *The three events E, F, and G are said to be independent if*

$$P(EFG) = P(E)P(F)P(G)$$

$$P(EF) = P(E)P(F)$$

$$P(EG) = P(E)P(G)$$

$$P(FG) = P(F)P(G)$$

Random Variables and Expectation

Random Variables

Definition. *Quantities determined by the result of an experiment. They have probabilityy distributions.*

Types of Random Variables

Discrete

Definition. *For a discrete random variable X, its probability mass function is*

$$p(a) = P\{X = a\}$$

Additionally,

$$p(x) \geq 0$$

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Definition. *The cumulative distribution function F is*

$$F(a) = \sum_{x \leq a} p(x)$$

Continuous

Definition. *For a continuous random variable X, its probability density function is*

$$f : \mathbb{R} \to [0, \infty]$$

$$f(x) = P\{X = x\}$$

Additionally,

$$f(x) \geq 0 \forall x$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Definition. *The cumulative distribution function F is*

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{\infty} f(x)dx$$

Useful properties of CDFs:

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

$$P(X < x) = F_X(x) - P_X(x)$$

Jointly Distributed Random Variables

Let X, Y be random variables. The joint CDF of X and Y is

$$F(x, y) = P(X \leq x, Y \leq y), x, y \in \mathbb{R}$$

Note: X and Y can be both discrete or continuous.
If X, Y are both discrete, then the joint pmf is

$$p(x, y) = P(X = x, Y = y)$$

If X, Y are both continuous, then the joint pmf is

$$f(x, y) : \mathbb{R}^2 \to [0, \infty]$$

$$f(x, y) \geq 0 \forall x, y$$

$$\iint_{(x,y) \in C} f(x, y)dxdy = P\{(X, Y) \in C\}$$

We call the (cdf, pmf, pdf) of X and Y the marginal (cdf, pmf, pdf).

Independent R.V.s

Independence
Two random variables X,Y are said to be independent if

$$P(X \leq x, Y \leq y) = P(X = x) \times P(Y = y) \forall x, y$$

If X and Y are discrete, X and Y are independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

If X and Y are continuous, X and Y are independent if

$$f(x, y) = f_X(x) f_Y(y)$$

Expectation

Expectation = E[X] = where does X take values? (location parameter)
X is discrete

$$E[X] = \sum_{\text{t possible values}} tP(X = t)$$

X is continuous

$$E[X] = \int_{-\infty}^{\infty} tf(t)dt$$

Properties of the Expected Value

Expected Value of Sums of Random Variables

Variance

Variance = Var(X) = how scattered are the values? (scale parameter)

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Covariance and Variance of Sums of R.V.s

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

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

Moment Generating Functions

$$\phi(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p() & if X is discrete \\ \int_{-\infty}^{infty} e^{tx} f(x) dx & if X is continuous \end{cases}$$