Probabilities and Statistics Refresher

Nayel Bettache

Department of Statistical Science, Cornell University

Sample space - The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

Sample space - The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

Event - Any subset E of the sample space is known as an event. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred.

Introduction to Probability and Combinatorics

Sample space - The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

Event - Any subset E of the sample space is known as an event. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred.

Axioms of probability - For each event E, we denote $P(E)$ as the probability of event E occuring. By noting $E_1, \ldots, E_n$ mutually exclusive events, we have the 3 following axioms:
(1) $0 \leqslant P(E) \leqslant 1$
(2) $P(S) = 1$
(3) $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P\left(E_i\right)$

Permutation - A permutation is an arrangement of r objects from a pool of n objects, in a given order. The number of such arrangements is given by $P(n, r)$, defined as:

$$P(n, r) = \frac{n!}{(n-r)!}.$$

Permutation - A permutation is an arrangement of r objects from a pool of n objects, in a given order. The number of such arrangements is given by $P(n, r)$, defined as:

$$P(n, r) = \frac{n!}{(n-r)!}.$$

Combination - A combination is an arrangement of r objects from a pool of n objects, where the order does not matter. The number of such arrangements is given by $C(n, r)$, defined as

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}.$$

Conditional Probability

Bayes' rule - For events A and B such that $P(B) > 0$, we have:

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

Conditional Probability

Bayes' rule - For events A and B such that $P(B) > 0$, we have:

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

Partition - Let $\{A_i, i \in \{1, \ldots, n\}\}$ be such that for all i, $A_i \neq \varnothing$. We say that $\{A_i\}$ is a partition if we have:

$$\forall i \neq j,\ A_i \cap A_j = \emptyset \quad \text{and} \quad \bigcup_{i=1}^{n} A_i = S$$

Conditional Probability

Bayes' rule - For events A and B such that $P(B) > 0$, we have:

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

Partition - Let $\{A_i, i \in \{1, \dots, n\}\}$ be such that for all $i, A_i \neq \varnothing$. We say that $\{A_i\}$ is a partition if we have:

$$\forall i \neq j,\ A_i \cap A_j = \emptyset \quad \text{and} \quad \bigcup_{i=1}^{n} A_i = S$$

Independence -Two events A and B are independent if and only if we have:

$$P(A \cap B) = P(A)P(B)$$

Random Variables

Random Variable - A random variable, often noted X, is a function that maps every element in a sample space to a real line.

Random Variables

Random Variable - A random variable, often noted X, is a function that maps every element in a sample space to a real line.

Cumulative distribution function (CDF) - The cumulative distribution function F which is monotonically non-decreasing and is such that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$, is defined as:

$$F(x) = P(X \leqslant x)$$

Random Variables

Random Variable - A random variable, often noted X, is a function that maps every element in a sample space to a real line.

Cumulative distribution function (CDF) - The cumulative distribution function F which is monotonically non-decreasing and is such that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$, is defined as:

$$F(x) = P(X \leqslant x)$$

Probability density function (PDF) - The probability density function f is the probability that X takes on values between two adjacent realizations of the random variable.

Random Variables

CDF and PDF - Properties in the discrete (D) and the continuous (C) cases:

Case	CDF F	PDF f	Properties of PDF
(D)	$F(x) = \sum_{x_i \leqslant x} P(X = x_i)$	$f(x_j) = P(X = x_j)$	$0 \leqslant f(x_j) \leqslant 1$ and $\sum_j f(x_j) = 1$
(C)	$F(x) = \int_{-\infty}^{x} f(y)dy$	$f(x) = \frac{dF}{dx}$	$f(x) \geqslant 0$ and $\int_{-\infty}^{+\infty} f(x)dx = 1$

Random Variables

Variance - The variance of a random variable, often noted Var(X) or σ^2, is a measure of the spread of its distribution function. It is determined as follows:

$$\text{Var}(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2.$$

Random Variables

Variance - The variance of a random variable, often noted Var(X) or σ^2, is a measure of the spread of its distribution function. It is determined as follows:

$$\text{Var}(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2.$$

Standard deviation - The standard deviation of a random variable, often noted σ, is a measure of the spread of its distribution function which is compatible with the units of the actual random variable. It is determined as follows:

$$\sigma = \sqrt{\text{Var}(X)}.$$

Random Variables

Variance - The variance of a random variable, often noted $\text{Var}(X)$ or σ^2, is a measure of the spread of its distribution function. It is determined as follows:

$$\text{Var}(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2.$$

Standard deviation - The standard deviation of a random variable, often noted σ, is a measure of the spread of its distribution function which is compatible with the units of the actual random variable. It is determined as follows:

$$\sigma = \sqrt{\text{Var}(X)}.$$

Expectation and Moments of the Distribution - Here are the expressions of the expected value $E[X]$, generalized expected value $E[g(X)]$, k^{th} moment $E\left[X^k\right]$ and characteristic function $\psi(\omega)$ for the discrete and continuous cases:

Case	$E[X]$	$E[g(X)]$	$E\left[X^k\right]$	$\psi(\omega)$
(D)	$\sum_{i=1}^{n} x_i f(x_i)$	$\sum_{i=1}^{n} g(x_i) f(x_i)$	$\sum_{i=1}^{n} x_i^k f(x_i)$	$\sum_{i=1}^{n} f(x_i) e^{i\omega x_i}$
(C)	$\int_{-\infty}^{+\infty} x f(x) dx$	$\int_{-\infty}^{+\infty} g(x) f(x) dx$	$\int_{-\infty}^{+\infty} x^k f(x) dx$	$\int_{-\infty}^{+\infty} f(x) e^{i\omega x} dx$

Jointly Distributed Random Variables

Conditional density - The conditional density of X w.r.t. Y, noted $f_{X|Y}$, is:

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

Jointly Distributed Random Variables

Conditional density - The conditional density of X w.r.t. Y, noted $f_{X|Y}$, is:

$$f_{X|Y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}.$$

Independence - Two random variables X and Y are said to be independent if:

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

Jointly Distributed Random Variables

Conditional density - The conditional density of X w.r.t. Y, noted $f_{X|Y}$, is:

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

Independence - Two random variables X and Y are said to be independent if:

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

Covariance - We define the covariance of two random variables X and Y, that we note σ_{XY}^2 or more commonly $\mathrm{Cov}(X, Y)$, as follows:

$$\mathrm{Cov}(X, Y) \triangleq \sigma_{XY}^2 = E\left[(X - \mu_X)(Y - \mu_Y)\right] = E[XY] - \mu_X \mu_Y$$

Jointly Distributed Random Variables

Conditional density - The conditional density of X w.r.t. Y, noted $f_{X|Y}$, is:

$$f_{X|Y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}.$$

Independence - Two random variables X and Y are said to be independent if:

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

Covariance - We define the covariance of two random variables X and Y, that we note σ_{XY}^2 or more commonly $\mathrm{Cov}(X,Y)$, as follows:

$$\mathrm{Cov}(X,Y) \triangleq \sigma_{XY}^2 = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E[XY] - \mu_X\mu_Y$$

Correlation - By noting σ_X, σ_Y the standard deviations of X and Y, we define the correlation between the random variables X and Y, noted ρ_{XY}, as follows:

$$\rho_{XY} = \frac{\sigma_{XY}^2}{\sigma_X\sigma_Y}$$

Parameter estimation

Random sample - A random sample is a collection of n random variables $X_1, \ldots, X_n$ that are independent and identically distributed with X.

Parameter estimation

Random sample - A random sample is a collection of n random variables $X_1, \ldots, X_n$ that are independent and identically distributed with X.

Estimator - An estimator $\hat{\theta}$ is a function of the data that is used to infer the value of an unknown parameter θ^* in a statistical model.

Parameter estimation

Random sample - A random sample is a collection of n random variables $X_1, \ldots, X_n$ that are independent and identically distributed with X.

Estimator - An estimator $\hat{\theta}$ is a function of the data that is used to infer the value of an unknown parameter θ^* in a statistical model.

Bias - The bias of an estimator $\hat{\theta}$ is defined as being the difference between the expected value of the distribution of $\hat{\theta}$ and the value of the true parameter, i.e.:

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta^*.$$

Parameter estimation

Random sample - A random sample is a collection of n random variables $X_1, \dots, X_n$ that are independent and identically distributed with X.

Estimator - An estimator $\hat{\theta}$ is a function of the data that is used to infer the value of an unknown parameter θ^* in a statistical model.

Bias - The bias of an estimator $\hat{\theta}$ is defined as being the difference between the expected value of the distribution of $\hat{\theta}$ and the value of the true parameter, i.e.:

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta^*.$$

Sample mean and variance - The sample mean and the sample variance of a random sample are used to estimate the true mean μ and the true variance σ^2 of a distribution, are noted $\bar{X}$ and s^2 respectively, and are such that:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad s^2 = \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$