

Contents

1	Definitions and Theory	1
1.1	Fundamentals	1
1.1.1	Expectation, Variance and Moments	2
1.1.2	Distributions	4
1.2	Sampling Theory	7

1

Definitions and Theory

1.1 Fundamentals

Definition 1: Two random vectors (X_1, X_2, \dots, X_n) and (Y_1, Y_2, \dots, Y_n) are said to be **independent** if $F(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = F_1(x_1, x_2, \dots, x_m)F_2(y_1, y_2, \dots, y_n)$ for all $(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) \in \mathbb{R}^{m+n}$ where F, F_1, F_2 are the joint CDF's of $(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n)$, (X_1, X_2, \dots, X_m) and (Y_1, Y_2, \dots, Y_n) respectively.

Theorem 1: Let $X = (X_1, X_2, \dots, X_m)$ and $Y = (Y_1, Y_2, \dots, Y_n)$ be independent random vectors. Then the component X_j of X ($j = 1, 2, \dots, m$) and the component Y_k of Y ($k = 1, 2, \dots, n$) are independent random variables. If h and g are Borel-measurable functions, $h(X_1, X_2, \dots, X_m)$ and $g(Y_1, Y_2, \dots, Y_n)$ are independent.

Theorem 2: Let X be a random variable with pdf $f(x)$. Then the pdf of $aX + b$ where $a \neq 0, b \in \mathbb{R}$ is given by $\frac{1}{|a|}f\left(\frac{x-b}{a}\right)$

Theorem 3: If X and Y are independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is $f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw$

1.1.1 Expectation, Variance and Moments

Definition 2: The **expectation** of a discrete random variable X having values x_1, x_2, \dots, x_n and probability function $f(x)$ is defined as $E(X) = \sum_{i=1}^n x_i f(x_i)$.

If X is a discrete random variable taking on infinite set of values x_1, x_2, \dots , then $E(X) = \sum_{i=1}^{\infty} x_i f(x_i)$ provided the infinite series converges absolutely.

For a continuous random variable X with distribution function $f(x)$, the expectation of X is defined as $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ provided the integral converges absolutely.

Theorem 4: The expectation has the following properties:

1. $E(cX) = cE(X)$ where c is any constant
2. If X and Y are any random variables then $E(X + Y) = E(X) + E(Y)$
3. If X and Y are independent random variables, then $E(XY) = E(X)E(Y)$

Definition 3: The **variance** is defined as $\text{Var}(X) = \sigma^2 = E[(X - \mu)^2]$. The **standard deviation** is defined as $\sigma = \sqrt{\text{Var}(X)}$.

Theorem 5: The variance has the following properties:

1. $\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$ where $\mu = E(X)$
2. If c is any constant, $\text{Var}(cX) = c^2 \text{Var}(X)$
3. The quantity $E[(X - a)^2]$ is a minimum where $a = \mu = E(X)$
4. If X and Y are independent random variables, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ and $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$

Definition 4: The r -th **moment** of a random variable X about the mean μ is defined as $\mu_r = E[(X - \mu)^r]$ where $r = 0, 1, 2, \dots$

The r -th moment of X about the origin is defined as $\mu'_r = E(X^r)$.

It follows that $\mu_0 = 0$, $\mu_1 = 1$, $\mu_2 = \sigma^2$

Theorem 6 (Law of the unconscious statistician - LOTUS): Let X be a discrete random variable with probability function $f(x)$. Then $Y = g(x)$ is also a discrete random variable.

The probability function of Y is $h(y) = P(Y = y) = \sum_{\{x|g(x)=y\}} P(X = x) = \sum_{\{x|g(x)=y\}} f(x)$.

If X takes on values x_1, x_2, \dots, x_n and Y takes on values y_1, y_2, \dots, y_m , then $m \leq n$ and $y_1h(y_1) + y_2h(y_2) + \dots + y_mh(y_m) = g(x_1)f(x_1) + g(x_2)f(x_2) + \dots + g(x_n)f(x_n)$ which lets us write the expectation of Y as

$$E(Y) = g(x_1)f(x_1) + g(x_2)f(x_2) + \dots + g(x_n)f(x_n) = \sum_{i=1}^n g(x_i)f(x_i).$$

Similarly when X is a continuous random variable and $Y = g(X)$, then $E(Y) = \int_{-\infty}^{\infty} g(x)f(x)dx$.

Definition 5: Let X be a random variable defined on (Ω, \mathcal{F}, P) . The function $M(t) = E[e^{tX}]$ is called the **moment generating function** (MGF) of the random variable X if the expectation on the right side exists in some neighbourhood of the origin. If the expectation on the right side does not exist in any neighbourhood of the origin, then we say the MGF does not exist.

The r -th derivative of the moment generating function is the r -th moment about the origin μ'_r .

Theorem 7: If the MGF $M(s)$ of a random variable X exists, then the MGF $M(s)$ has derivatives of all orders at $s = 0$ and

$$M^{(k)}(s)|_{s=0} = EX^k \text{ for positive integer } k$$

Theorem 8: The moment generating function has the following properties:

1. For any constants a and b , the mgf of the random variable $aX + b$ is given by $M_{aX+b} = e^{bt}M_X(at)$
2. If X and Y are independent random variables having moment generating functions $M_X(t)$ and $M_Y(t)$ respectively, then $M_{X+Y}(t) = M_X(t)M_Y(t)$
3. **Uniqueness Theorem** Suppose that X and Y are random variables having moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. Then X and Y have the same probability distribution if and only if $M_X(t) = M_Y(t)$ identically.

Definition 6: Let X_1, X_2, \dots, X_n be a jointly distributed or (X_1, X_2, \dots, X_n) be a random vector. If $E[\exp(\sum_{j=1}^n t_j X_j)]$ exists for $|t_j| \leq h_j, j = 1, 2, \dots, n$, we write $M(t_1, t_2, \dots, t_n) = E[\exp(t_1 X_1 + t_2 X_2 + \dots + t_n X_n)]$ and call it the MGF of X_1, X_2, \dots, X_n or simply, the **joint moment generating function** (joint MGF) of the random vector (X_1, X_2, \dots, X_n)

Theorem 9: The joint MGF $M(t_1, t_2)$ uniquely determines the joint distribution of (X, Y) . Conversely, if the joint MGF exists it is unique.

Theorem 10: The joint MGF $M(t_1, t_2)$ completely determines the marginal distributions of X and Y .

$$M(t_1, 0) = E[\exp(t_1 X)] = M_X(t_1) \text{ and } M(0, t_2) = E[\exp(t_2 X)] = M_Y(t_2)$$

Theorem 11: X and Y are independent random variables if and only if $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$ for all $t_1 \in [-h_1, h_1], t_2 \in [-h_2, h_2]$

1.1.2 Distributions

Definition 7: Let p be the probability that an event will happen in any single Bernoulli trial (trial with outcomes either success or failure). The probability that an event will happen exactly x times in n trials is given by the **Binomial Random Variable** with pmf distribution $f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$, where $x = 0, 1, \dots, n$.

Proposition 1.1.3: The binomial random variable has mean $\mu = np$ and variance $\sigma^2 = npq$.

Definition 8: The **Poisson Random Variable** has pmf distribution $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, \dots$ and $\lambda > 0$.

Proposition 1.1.4: The Poisson Random Variable has mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$.

Proposition 1.1.5: The Poisson Random Variable with $\lambda = np$ is the limiting case of the Binomial Distribution. It approximates the Binomial Random variable $\text{Binomial}(n, p)$ when n is large and probability of occurrence of an event p is close to 0.

Definition 9: The **uniform random variable** has the pdf distribution $f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$.

Proposition 1.1.6: The uniform random variable has mean $\mu = \frac{a+b}{2}$ and variance $\sigma^2 = \frac{1}{12}(b-a)^2$.

Definition 10: The **exponential random variable** has the pdf distribution $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$

Proposition 1.1.7: The exponential random variable has mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$.

Definition 11: The **normal random variable**, also known as the gaussian random variable, has pdf distribution $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ where $-\infty < x < \infty$ where μ and σ are the mean and standard deviation respectively.

Definition 12: When $\mu = 0$ and $\sigma = 1$, we get the **standard normal random variable** with distribution $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$.

We can write any normal random variable $Y \sim N(\mu, \sigma)$ in terms of the standard normal random variable $X \sim N(0, 1)$ as $Y = \sigma X + \mu$.

Proposition 1.1.8: When n is large and neither p or q is too close to 0, the binomial random variable X can be approximated by a normal distribution with mean np and standard deviation \sqrt{npq} . The approximation is very good when $np, nq > 5$.

Proposition 1.1.9: The Poisson Distribution approaches the normal distribution $N(\lambda, \sqrt{\lambda})$ as $\lambda \rightarrow \infty$, i.e. the Poisson distribution is asymptotically normal.

Definition 13: A variable the pdf $f(x) = \frac{1}{\sigma\pi(1 + (\frac{x-\mu}{\sigma})^2)}$, $x \in \mathbb{R}$ where $\sigma > 0, \mu \in \mathbb{R}$ is called a **Cauchy random variable** with parameter μ and σ^2 . We write $X \sim \mathcal{C}(\mu, \sigma^2)$ for Cauchy random X with pdf.

Definition 14: When $\mu = 0$ and $\sigma^2 = 1$, we get the **standard Cauchy random variable** $C(0, 1)$ with distribution $f(x) = \frac{1}{\pi(1+x^2)}$.

We can write any Cauchy random variable $Y = C(\mu, \sigma^2)$ in terms of the standard Cauchy random variable $X = C(0, 1)$ as $Y = \sigma X + \mu$.

Proposition 1.1.10: The mean, variance, higher moments, moment generating function of a Cauchy random variable do not exist.

Definition 15: $\Gamma(\alpha)$ denotes the gamma function which is defined as $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$, where $\alpha > 0$. It is a commonly used extension of the factorial function.

Theorem 12: The Gamma function has the following properties:

1. The improper integral converges for all $\alpha > 0$.
2. $\Gamma(\alpha) > 0$ for all $\alpha > 0$, $\Gamma(\alpha) \rightarrow \infty$ as $\alpha \rightarrow 0$.
3. $\Gamma(1) = 1$
4. $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$
5. For $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$
6. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Definition 16: A random variable with the pdf $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$ where $0 < x < \infty, \alpha > 0, \beta > 0$ is called a **gamma random variable** $G(\alpha, \beta)$ with parameters α and β .

Proposition 1.1.11: When $\alpha = 1$, we see that the gamma distribution is a generalization of the exponential distribution as $G(1, \beta) = \exp(\beta)$ with pdf $f(x) = \frac{1}{\beta} e^{-x/\beta}$, $x > 0$.

Proposition 1.1.12: The gamma distribution has mean $\mu = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$.

Definition 17: The **chi-square distribution** is

$$f(x) = \frac{1}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-x/2} \text{ where } 0 < x < \infty.$$

χ_p^2 denotes a chi-square random variable with p degrees of freedom.

The chi-square distribution is a special case of the gamma distribution with $\alpha = \frac{p}{2}$ and $\beta = 2$.

Proposition 1.1.13: The mean of the chi-square distribution is given by $\mu = p$ and the variance is given by $\sigma^2 = 2p$.

Proposition 1.1.14: Let $X \sim N(0, 1)$. Then $X^2 \sim \chi_1^2$

Proposition 1.1.15: Let X_1, X_2, \dots, X_n be independent normal random variables with mean 0 and variance 1. Then $\chi^2 = X_1^2 + X_2^2 + \dots + X_p^2$ is chi-square distributed with p degrees of freedom.

1.2 Sampling Theory

We can either sample *with replacement* or *without replacement*. A finite population sampled with replacement can be considered infinite. Sampling from a very large finite population can similarly be considered as sampling from an infinite population.

To properly choose the sample, we can make sure that every member of the population has an equal chance of being in the sample. Normally, since the sample size is much smaller than the population size, sampling without replacement will give practically the same results as sampling with replacement.

For a sample of size n from a population which we assume has distribution $f(x)$, we can choose members of the population at random, each selection corresponding to a random variable X_1, X_2, \dots, X_n with corresponding values x_1, x_2, \dots, x_n . In case we are assuming sampling without replacement, X_1, X_2, \dots, X_n will be independent and identically distributed random variables with probability distribution $f(x)$.

Definition 18: Let X be a random variable with a distribution f , and let X_1, X_2, \dots, X_n be iid random variables with the common distribution f .

Then the collection X_1, X_2, \dots, X_n is called a **random sample** of size n from the population f .

Since X_1, X_2, \dots, X_n are iid, the joint distribution of the random sample is $f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2)\dots f(x_n)$.

Any quantity obtained from a sample for the purpose of estimating a population parameter is called a sample statistic, or briefly statistic. Mathematically, a sample statistic for a sample of size n can be defined as a function of the random variables X_1, X_2, \dots, X_n as $T(X_1, X_2, \dots, X_n)$. This itself is a random variable whose values can be represented as $T(x_1, x_2, \dots, x_n)$.

Definition 19: Let X_1, X_2, \dots, X_n be a random sample of size n from the population whose distribution is $f(x|\theta)$ (the distribution f with unknown parameter θ). Let $T(x_1, x_2, \dots, x_n)$ be a real-valued or vector-valued function whose domain includes the range of (X_1, X_2, \dots, X_n) . Then the random variable or random vector $Y = T(X_1, \dots, X_n)$ is called a **statistic** provided that T is not a function of any unknown parameter θ .

For example consider $X \approx N(\mu, \sigma^2)$ where μ is known but σ is unknown. Then $\frac{\sum_{i=1}^n X_i}{\sigma^2}$ is not a statistic but $\frac{\sum_{i=1}^n X_i}{\mu^2}$ is a statistic.

Two common statistics are the sample mean and sample variance.

Definition 20: The **sample mean** is the arithmetic average of the values in the random sample. It is denoted by $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$.

The **sample variance** is the statistic defined by $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$

The **sample standard deviation** is the statistic defined by $S = \sqrt{S^2}$.

Definition 21: Let X_1, X_2, \dots, X_n be a random sample from a population $f(x|\theta)$. We say that a statistic $T(X_1, X_2, \dots, X_n)$ is an **unbiased estimator** of the parameter θ if $E(T) = \theta$ for all possible values of θ .

Theorem 13: Let X_1, X_2, \dots, X_n be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then:

1. $E(\bar{X}) = \mu$
2. $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$
3. $E(S^2) = \sigma^2$

From the above theorem we see that the sample mean \bar{X} is an unbiased estimator of the population mean μ and the sample variance S^2 is an unbiased estimator of the population variance σ^2 . (The reason we included $1/n - 1$ in the definition of the sample variance was to make it an unbiased estimator)

Definition 22: Let X_1, X_2, \dots, X_n be a random sample of size n from a population $f(x|\theta)$. The probability distribution of a statistic $T(X_1, X_2, \dots, X_n)$ is called the sampling distribution of T .

Theorem 14 (Distribution of the sample mean): Let X_1, X_2, \dots, X_n be a random sample from a population with MGF $M_X(t)$. Then the MGF of the sample mean is $M_{\bar{X}}(t) = (M_X(t/n))^n$.

Theorem 15: Let X_1, X_2, \dots, X_n be a random sample from a $N(\mu, \sigma^2)$ distribution, and let \bar{X} denote the sample mean.

Then \bar{X} and the random vector $(X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ are independent.

Theorem 16: Let X_1, X_2, \dots, X_n be a random sample from a $N(\mu, \sigma^2)$ distribution, and let \bar{X} denote the sample mean and S^2 denote the sample variance. Then \bar{X} and S^2 are independent random variables.

The converse of this theorem is also true: if the sample mean and sample variance of a random sample are independent random variables then population distribution is normal.